INTRODUÇÃO À GENÉTICA

Anthony J. F. Griffiths

Susan R. Wessler

Sean B. Carroll

John Doebley

11ª EDIÇÃO

gen | GUANABARA KOOGAN

INTRODUÇÃO À
GENÉTICA

O GEN | Grupo Editorial Nacional, a maior plataforma editorial no segmento CTP (científico, técnico e profissional), publica nas áreas de saúde, ciências exatas, jurídicas, sociais aplicadas, humanas e de concursos, além de prover serviços direcionados a educação, capacitação médica continuada e preparação para concursos. Conheça nosso catálogo, composto por mais de cinco mil obras e três mil e-books, em www.grupogen.com.br.

As editoras que integram o GEN, respeitadas no mercado editorial, construíram catálogos inigualáveis, com obras decisivas na formação acadêmica e no aperfeiçoamento de várias gerações de profissionais e de estudantes de Administração, Direito, Engenharia, Enfermagem, Fisioterapia, Medicina, Odontologia, Educação Física e muitas outras ciências, tendo se tornado sinônimo de seriedade e respeito.

Nossa missão é prover o melhor conteúdo científico e distribuí-lo de maneira flexível e conveniente, a preços justos, gerando benefícios e servindo a autores, docentes, livreiros, funcionários, colaboradores e acionistas.

Nosso comportamento ético incondicional e nossa responsabilidade social e ambiental são reforçados pela natureza educacional de nossa atividade, sem comprometer o crescimento contínuo e a rentabilidade do grupo.

INTRODUÇÃO À GENÉTICA

Anthony J. F. Griffiths
University of British Columbia

Susan R. Wessler
University of California, Riverside

Sean B. Carroll
Howard Hughes Medical Institute
University of Wisconsin-Madison

John Doebley
University of Wisconsin-Madison

Revisão Técnica

Márcia Mattos Gonçalves Pimentel

Bacharel em Ciências Biológicas. Mestre em Ciências Biológicas (Genética Humana) pela Universidade Federal do Rio de Janeiro (UFRJ). Doutora em Ciências (Genética Humana) pela UFRJ. Professora Associada do Departamento de Genética do Instituto de Biologia Roberto Alcântara Gomes da Universidade do Estado do Rio de Janeiro (UERJ). Coordenadora do Serviço de Genética Humana da UERJ.

Tradução

Sylvia Elgg Roberto

Décima primeira edição

gen | GUANABARA KOOGAN

- Os autores deste livro e a EDITORA GUANABARA KOOGAN LTDA. empenharam seus melhores esforços para assegurar que as informações e os procedimentos apresentados no texto estejam em acordo com os padrões aceitos à época da publicação, *e todos os dados foram atualizados pelos autores até a data da entrega dos originais à editora*. Entretanto, tendo em conta a evolução das ciências da saúde, as mudanças regulamentares governamentais e o constante fluxo de novas informações sobre terapêutica medicamentosa e reações adversas a fármacos, recomendamos enfaticamente que os leitores consultem sempre outras fontes fidedignas, de modo a se certificarem de que as informações contidas neste livro estão corretas e de que não houve alterações nas dosagens recomendadas ou na legislação regulamentadora. *Adicionalmente, os leitores podem buscar por possíveis atualizações da obra em http://gen-io.grupogen.com.br*.

- Os autores e a editora se empenharam para citar adequadamente e dar o devido crédito a todos os detentores de direitos autorais de qualquer material utilizado neste livro, dispondo-se a possíveis acertos posteriores caso, inadvertida e involuntariamente, a identificação de algum deles tenha sido omitida.

- INTRODUCTION TO GENETIC ANALYSIS, ELEVENTH EDITION
  First published in the United States by
  W.H. FREEMAN AND COMPANY, New York
  Copyright © 2015 W.H. FREEMAN AND COMPANY.
  All Rights Reserved.

- Publicado originalmente nos Estados Unidos por
  W.H. FREEMAN AND COMPANY, New York
  Copyright © 2015 W.H. FREEMAN AND COMPANY.
  Todos os Direitos Reservados.
  ISBN: 978-1-4641-0948-5

- Direitos exclusivos para a língua portuguesa
  Copyright © 2016 by
  EDITORA GUANABARA KOOGAN LTDA.
  Uma editora integrante do GEN | Grupo Editorial Nacional
  Travessa do Ouvidor, 11
  Rio de Janeiro – RJ – CEP 20040-040
  Tels.: (21) 3543-0770/(11) 5080-0770 | Fax: (21) 3543-0896
  www.grupogen.com.br | editorial.saude@grupogen.com.br

- Reservados todos os direitos. É proibida a duplicação ou reprodução deste volume, no todo ou em parte, em quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição pela Internet ou outros), sem permissão, por escrito, da EDITORA GUANABARA KOOGAN LTDA.

- Capa: Vicki Tomaselli
  Produção digital: Geethik

- Ficha catalográfica

I48

11. ed.

Introdução à genética/Anthony J. F. Griffiths ...[et. al]; tradução Sylvia Werdmüller von Elgg Roberto. – 11. ed. – Rio de Janeiro: Guanabarra Koogan, 2016.

il.

Tradução de: Introduction to genetic analysis

ISBN 978-85-277-2995-6

1. Genética. 2. Biologia. I. Griffiths, Anthony J. F. . II. Roberto, Sylvia Werdmüller von Elgg.

16-33012

CDD: 576.5
CDU: 575

Sobre os autores

[*Barbara Moon.]*

Anthony J. F. Griffiths é Professor Emérito de Botânica na University of British Columbia. Sua pesquisa é voltada para a genética do desenvolvimento utilizando como modelo o fungo *Neurospora crassa*. Foi Presidente da Genetics Society of Canada e Secretário-Geral da International Genetics Federation. Foi recentemente laureado com a Fellow Medal, da International Mycological Association.

[*Iqbal Pittawala.]*

Susan R. Wessler é Professora especializada em Genética no Department

of Botany and Plant Sciences, da University of California, Riverside. Sua pesquisa é voltada para os elementos de transposição de plantas e sua contribuição para a evolução dos genomas e dos genes. Dra. Wessler foi eleita para a National Academy of Sciences em 1998. Como Professora do Howard Hughes Medical Institute, desenvolveu e ministra aulas em uma série de cursos sobre genoma dinâmico, nas quais graduandos vivenciam a interessante experiência da descoberta científica.

[*Sean Carroll.]*

Sean B. Carroll é Vice-presidente de Educação Científica no Howard Hughes Medical Institute e Professor de Biologia Molecular e Genética na University of Wisconsin–Madison. Dr. Carroll é um líder no campo da biologia do desenvolvimento evolutivo e foi eleito para a National Academy of Sciences em 2007. Ele também é o autor de *Brave Genius*, *Endless Forms Most Beautiful: The Making of the Fittest* e *Remarkable Creatures*, finalista do National Book Award na categoria não ficção em 2009.

[*John Doebley.]*

John Doebley é Professor de Genética na University of Wisconsin–Madison. Ele estuda a genética da domesticação de plantas cultivadas usando os métodos de genética de populações e quantitativa. Em 2003, foi eleito para a National Academy of Sciences e, em 2005, foi Presidente da American Genetics Association. Ensina Genética Geral e Evolutiva na University of Wisconsin.

Prefácio

Desde a primeira edição, publicada em 1974, *Introdução à Genética* enfatiza a força e a eficácia da abordagem genética na pesquisa biológica e suas aplicações. Em suas muitas edições, ampliou-se continuamente a abrangência do texto, à medida que a força da análise genética tradicional se estendeu com a introdução da tecnologia do DNA recombinante e em seguida da genômica. Na 11ª edição, damos continuidade a essa tradição e mostramos de que maneira o surgimento desse tipo de análise foi inspirador nas pesquisas em biologia, agricultura e saúde humana.

Ferramentas pedagógicas

Uma das novidades mais importantes desta nova edição é a inclusão de uma lista de **resultados de aprendizagem** no início de cada capítulo. Os resultados de aprendizagem são elementos cruciais de entendimento. Um dos princípios da teoria construtivista do aprendizado versa sobre o fato de que, embora o entendimento possa ser uma série de novos circuitos mentais, o estudante nunca é capaz de saber o que se encontra em seu cérebro até que busque realizar com tal conhecimento algum tipo de atividade. De fato, o entendimento já foi até mesmo definido por alguns como *capacidade flexível de realização*. A lista de objetivos mostra aos estudantes quais realizações específicas são esperadas. As notas a seguir mostram como os benefícios dos resultados de aprendizagem presentes neste livro podem ser ampliados por professores que queiram utilizá-los.

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Realizar uma análise quantitativa da progênie de um cruzamento-teste di-híbrido para avaliar se os dois genes estão ligados no mesmo cromossomo

ou não
- Estender o mesmo tipo de análise a diversos *loci* para produzir um mapa das posições relativas dos *loci* em um cromossomo
- Em fungos ascomicetos, mapear os centrômeros a outros *loci* ligados
- Em ascos, prever as proporções de alelos oriundos de etapas específicas no modelo heterodúplex de *crossing over*.

Aulas ministradas para turmas grandes ou pequenas (por exemplo, palestras e tutoriais) devem ser estruturadas, o máximo possível, sobre resultados de aprendizagem semelhantes aos apresentados nos capítulos. Em vários momentos, os alunos presentes devem ser solicitados a demonstrar o quanto compreenderam o material trabalhado por meio da tentativa de realização de um ou mais dos resultados. Em provas discursivas ou testes, o professor deve tentar ater-se ao máximo aos resultados de aprendizagem. Ao corrigir os testes, mostrar de que maneira os resultados foram ou não atingidos pelo aprendiz.

Os estudantes devem ler a lista de resultados de aprendizagem antes de iniciarem um capítulo. Embora não seja possível compreender a maior parte deles antes da leitura, seu texto passa uma ideia do que se encontra à frente, mostrando quais são as expectativas do professor. É ideal que, após a leitura da seção de um capítulo, os alunos voltem à lista e encontrem a que resultado se aplica o material recém-estudado, processo que deve ser repetido também ao fim do capítulo, para os alunos fazerem uma varredura nas seções e ligarem cada assunto a seu respectivo resultado esperado tanto quanto possível. Durante a resolução dos problemas finais de cada capítulo, os alunos devem focar seus esforços nas habilidades descritas nos resultados de aprendizagem, além de utilizá-los como revisão rápida ao estudar para provas, tentando imaginar de que maneiras será esperado que demonstrem entendimento por meio da aplicação de tais resultados.

O objetivo geral de um curso de genética é aprender a pensar e trabalhar como um geneticista. Os resultados de aprendizagem dividem tal objetivo geral nas diversas habilidades diferentes exigidas por esse assunto analítico.

Nesta edição, substituímos as “Mensagens” por “**Conceitos-chave**”. As mensagens fizeram parte desta obra desde sua primeira edição, em 1974. Nos anos 1960 e 70, talvez em virtude da popularidade do princípio de Marshall McLuhan, “O meio é a mensagem”, a palavra *mensagem* tinha um uso comum, e professores eram com frequência indagados: “Qual é a sua mensagem?” Embora, com a ascensão da mídia eletrônica, talvez seja o momento de trazer de volta o princípio de McLuhan, sentimos que a palavra *mensagem* já não tem o mesmo significado que carregava em 1974.

Nova cobertura da análise genética moderna

Um dos objetivos esperados com a leitura desta obra é mostrar como a identificação de genes e suas interações é um recurso poderoso para se entender as propriedades biológicas. Nesta 11ª edição, o Capítulo 1, introdutório, foi completamente reescrito, focando nas aplicações modernas da genética. A partir desse capítulo, o estudante acompanha o processo de dissecção da genética tradicional, começando por uma cobertura em etapas da identificação de um gene no Capítulo 2, o mapeamento gênico no Capítulo 4 e a identificação de vias e redes pelas interações gênicas no Capítulo 6. Novas abordagens genômicas para identificar e localizar são exploradas nos Capítulos 10, 14 e 19.

O gene *SUB1* aumenta a produção do arroz sob alagamento

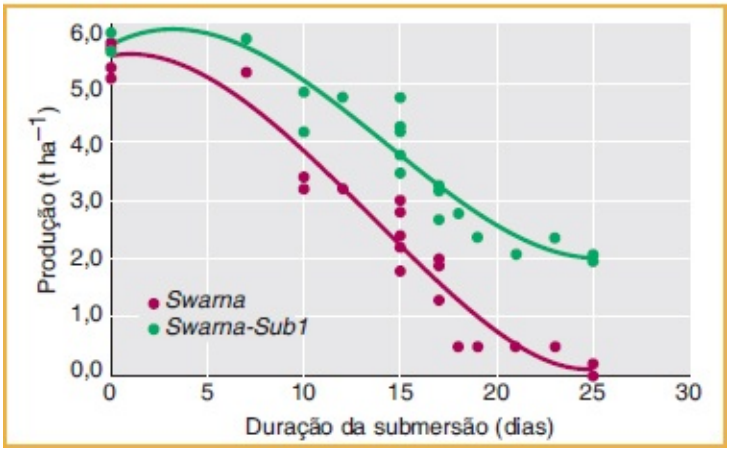


FIGURA 1.20 Comparação do rendimento da variedade *Swarna*, que não é tolerante ao alagamento (*círculos roxos*) e da variedade *Swarna-sub1*, que é tolerante (*círculos verdes*). Rendimento em toneladas por hectare (*eixo y*) *versus* duração do alagamento em dias (*eixo x*). (Dados de Ismail *et al.*, "The contribution of submergence-tolerant (Sub 1) rice varieties to food security in flood-prone rainfed lowland areas in Asia", *Field Crops Research* 152, 2013, 83-93, ©Elsevier.)

- O Capítulo 1, com novo conceito, visa despertar o interesse do estudante para a genética por meio da apresentação de uma seleção de aplicações modernas na biologia, na evolução, na medicina e na agricultura. Após uma breve história do estudo de genética e uma revisão sobre seus fundamentos, o capítulo descreve quatro utilizações da genética nos dias de hoje
- Os Capítulos 2 e 4 apresentam uma dissecção da genética clássica com uma introdução mais gradual. O Capítulo 2 inicia com uma nova introdução à genética e ao papel da análise genética na identificação de traços de herança monogênica. Os cruzamentos são mostrados visual e matematicamente, e os conceitos de dominância e recessividade são explicados em termos de haplossuficiência e haploinsuficiência. Os usos do teste do qui-quadrado, no Capítulo 4, foram reescritos com mais clareza
- A aplicação moderna da genética, introduzida no Capítulo 1, tem prosseguimento no Capítulo 14, por meio da aplicação de novas técnicas genômicas, como os sequenciamentos de RNA e exomas, os quais são utilizados na resolução de problemas em medicina. A busca por significado nos segmentos não codificadores do genoma é uma fronteira importante na genômica, e o projeto ENCODE foi adicionado ao capítulo em questão como exemplo dessa pesquisa.

Foco nos avanços mais importantes da genética

Aprimoramos a abordagem a diversos tópicos importantes nesta 11ª edição.

Remodelagem da cromatina e epigenética. Antes desmembrado em vários capítulos, o nascente campo da epigenética consta por completo no Capítulo 12. Na seção 12.3, "Cromatina dinâmica", tratamos dos três maiores mecanismos de alteração da estrutura da cromatina: remodelagem de cromatina, modificação das

histonas e variantes de histonas. Foram feitas modificações ao longo de toda a seção a fim de detalhar e esclarecer melhor o assunto, com base em avanços recentes no campo em questão.

Vigilância do genoma. Pesquisas de ponta sobre os elementos de transposição revelaram sistemas de vigilância do genoma em plantas, animais e bactérias semelhantes ao anteriormente identificado em *C. elegans*. O Capítulo 15 traz uma visão geral sobre piRNA em animais e crRNA em bactérias, além de permitir que os alunos comparem e contrastem essas abordagens com elementos Tc1 em nematódeos e MITE em plantas.

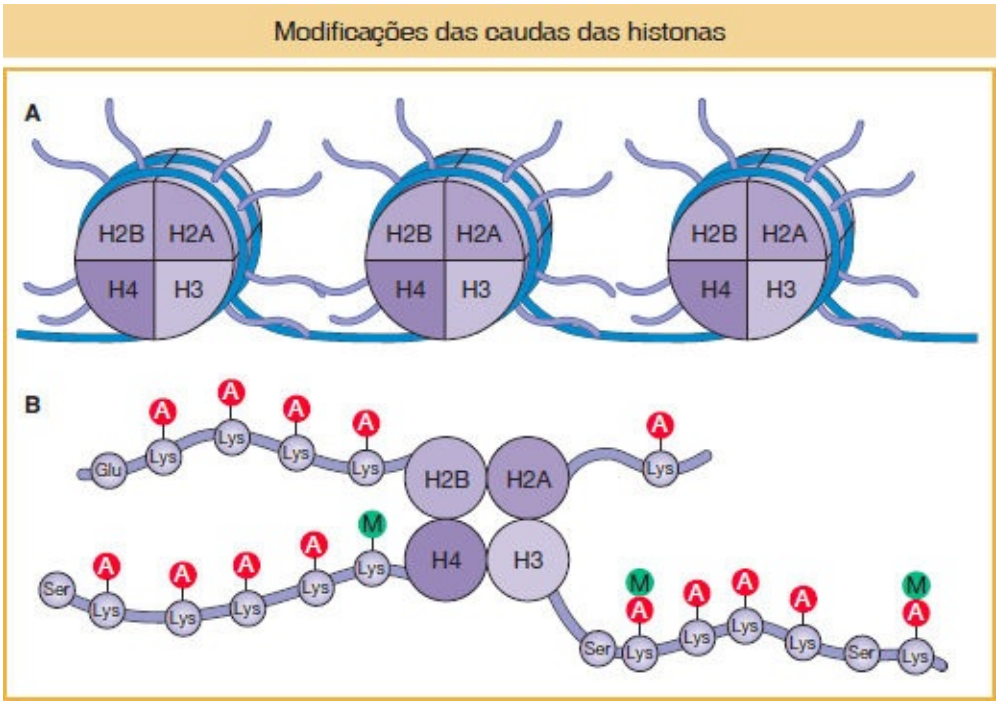


Figura 12.13 A. As caudas das histonas se projetam a partir do cerne do nucleossomo (*vermelho*). **B.** Estão demonstrados exemplos de modificações nas caudas das histonas. Os círculos com A representam acetilação, enquanto os círculos com M representam metilação. Ver o texto para os detalhes.

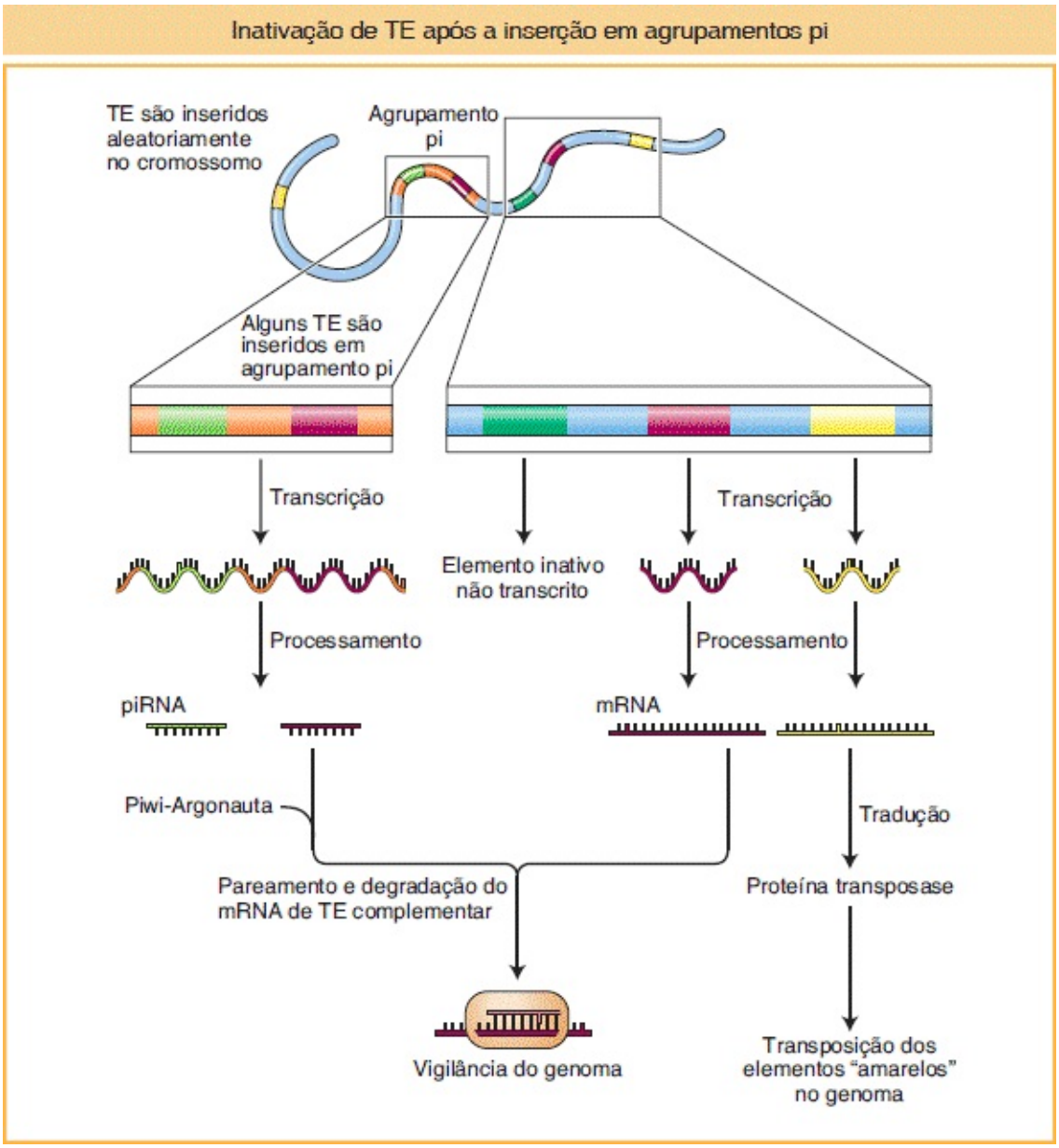


Figura 15.27 A inserção dos transpósons *verde* e *cor-de-rosa* em um agrupamento pi no genoma resulta na degradação de transcritos destes dois transpósons por meio das etapas demonstradas e descritas no texto. Contrariamente, o transpóson *amarelo* permanecerá ativo até que cópias sejam inseridas ao acaso em um agrupamento pi.

Aspectos mantidos

Cobertura de organismos-modelo

Esta 11ª edição mantém a cobertura a respeito dos sistemas que servem de

modelo em formato prático e flexível tanto para estudantes como para professores.

- No Capítulo 1 há uma introdução a alguns dos principais organismos-modelo e esclarecimentos sobre o sucesso obtido pelo seu uso
- Os Boxes Organismo-modelo são apresentados no contexto apropriado, fornecendo informação adicional sobre o organismo na natureza e seu uso experimental
- Um Guia Resumido de Organismos-modelo, antes dos Apêndices, garante acesso rápido a informações essenciais e práticas sobre os usos de organismos-modelo específicos em pesquisas
- Um Índice de Organismos-modelo, no final do livro, cita as páginas do texto em que se encontram referências a organismos específicos, possibilitando aos professores e estudantes encontrar com facilidade e obter informação comparativa sobre os organismos.

Grupos de problemas

Não importa quão clara tenha sido a exposição, o entendimento profundo requer que o estudante se engaje pessoalmente com o material. Daí nosso esforço para estimulá-lo a resolver os problemas. Com o foco na análise genética, esta 11ª edição fornece aos estudantes oportunidades de praticar habilidades para resolver os problemas de acordo com o seguinte:

- **Conjuntos versáteis de problemas**: os problemas abrangem todos os graus de dificuldade, sendo categorizados de acordo com os níveis – básicos ou desafiadores
- **Questões sobre as figuras**: um novo grupo de problemas incluído no fim de cada capítulo abrange perguntas sobre as figuras estudadas. Essas questões estimulam o estudante a raciocinar sobre as figuras e o ajudam a avaliar seu entendimento dos principais conceitos
- **Problemas resolvidos**: no fim de cada capítulo, esses exemplos ilustram como os geneticistas aplicam os princípios aos dados experimentais

- **Como solucionar o problema**: um problema de genética é um esboço de uma matriz complexa de conceitos e informação. Esta parte ajuda os estudantes a aprenderem a usar uma abordagem estratégica para resolver o problema, uma etapa de cada vez, conceito por conceito.

Como a genética é praticada hoje

Uma seção intitulada “O que os geneticistas fazem atualmente” sugere como as técnicas genéticas estão sendo utilizadas atualmente para responder a questões biológicas específicas, como “Qual a ligação entre encurtamento do telômero e envelhecimento?” ou “Como podemos encontrar componentes perdidos em uma via biológica específica?”

Agradecimentos

Estendemos nossos agradecimentos e nossa gratidão aos colegas que fizeram a revisão desta edição e cujos conselhos e percepções foram muito úteis:

Anna Allen, *Howard University*
Melissa Antonio, *California Baptist University*
Dave Bachoon, *Georgia College & State University*
Brianne Barker, *Drew University*
Lina Begdache, *Binghamton University*
Edward Berger, *Dartmouth College*
Aimee Bernard, *University of Colorado Denver*
Jaime Blair, *Franklin & Marshall College*
Jay Brewster, *Pepperdine University*
Doug Broadfield, *Florida Atlantic University*
Mirjana Brockett, *Georgia Institute of Technology*
Judy Brusslan, *California State University, Long Beach*
Gerald Buldak, *Loyola University Chicago*
Aaron Cassill, *University of Texas at San Antonio*
Helen Chamberlin, *Ohio State University*
Henry Chang, *Purdue University*
Randolph Christensen, *Coe College*
Mary Clancy, *University of New Orleans*
Craig Coleman, *Brigham Young University*
Matthew Collier, *Wittenberg University*
Shannon Compton, *University of Massachusetts–Amherst*
Diane Cook, *Louisburg College*
Victoria Corbin, *University of Kansas*
Claudette Davis, *George Mason University*
Ann Marie Davison, *Kwantlen Polytechnic University*

Elizabeth De Stasio, *Lawrence University*
Matt Dean, *University of Southern California*
Michael Dohm, *Chaminade University*
Robert Dotson, *Tulane University*
Chunguang Du, *Montclair State University*
Erastus Dudley, *Huntingdon College*
Edward Eivers, *California State University, Los Angeles*
Robert Farrell, *Penn State University*
David Foltz, *Louisiana State University*
Wayne Forrester, *Indiana University*
Rachael French, *San Jose State University*
Shirlean Goodwin, *University of Memphis*
Topher Gee, *UNC Charlotte*
John Graham, *Berry College*
Theresa Grana, *University of Mary Washington*
Janet Guedon, *Duquesne University*
Patrick Gulick, *Concordia University*
Richard Heineman, *Kutztown University*
Anna Hicks, *Memorial University*
Susan Hoffman, *Miami University*
Stanton Hoegerman, *College of William and Mary*
Margaret Hollingsworth, *University at Buffalo*
Nancy Huang, *Colorado College*
Jeffrey Hughes, *Millikin University*
Varuni Jamburuthugoda, *Fordham University*
Pablo Jenik, *Franklin & Marshall College*
Aaron Johnson, *University of Colorado School of Medicine*
Anil Kapoor, *University of La Verne*
Jim Karagiannis, *University of Western Ontario*
Kathleen Karrer, *Marquette University*
Jessica Kaufman, *Endicott College*
Darrell Killian, *Colorado College*

Dennis Kraichely, *Cabrini College*
Anuj Kumar, *University of Michigan*
Janice Lai, *Austin Community College*
Evan Lau, *West Liberty University*
Min-Ken Liao, *Furman University*
Sarah Lijegren, *University of Mississippi*
Renyi Liu, *University of California, Riverside*
Diego Loayza, *Hunter College*
James Lodolce, *Loyola University Chicago*
Joshua Loomis, *Nova Southeastern University*
Amy Lyndaker, *Ithaca College*
Jessica Malisch, *Claremont McKenna College*
Patrick Martin, *North Carolina A&T State University*
Presley Martin, *Hamline University*
Dmitri Maslov, *University of California, Riverside*
Maria Julia Massimelli, *Claremont McKenna College*
Endre Mathe, *Vasile Goldis Western University of Arad*
Herman Mays, *University of Cincinnati*
Thomas McGuire, *Penn State Abington*
Mark Meade, *Jacksonville State University*
Ulrich Melcher, *Oklahoma State University*
Philip Meneely, *Haverford College*
Ron Michaelis, *Rutgers University*
Chris Mignone, *Berry College*
Sarah Mordan-McCombs, *Franklin College of Indiana*
Ann Murkowski, *North Seattle Community College*
Saraswathy Nair, *University of Texas at Brownsville*
Sang-Chul Nam, *Texas A&M International University*
Scot Nelson, *University of Hawaii at Manoa*
Brian Nichols, *University of Illinois at Chicago*
Todd Nickle, *Mount Royal University*
Juliet Noor, *Duke University*

Mohamed Noor, *Duke University*
Daniel Odom, *California State University, Northridge*
Kirk Olsen, *East Los Angeles College*
Kavita Oommen, *Georgia State University*
Maria Orive, *University of Kansas*
Laurie Pacarynuk, *University of Lethbridge*
Patricia Phelps, *Austin Community College*
Martin Poenie, *University of Texas at Austin*
Jennifer Powell, *Gettysburg College*
Robyn Puffenbarger, *Bridgewater College*
Jason Rauceo, *John Jay College (CUNY)*
Eugenia Ribiero-Hurley, *Fordham University*
Ronda Rolfes, *Georgetown University*
Edmund Rucker, *University of Kentucky*
Jeffrey Sands, *Lehigh University*
Monica Sauer, *University of Toronto at Scarborough, UTSC*
Ken Saville, *Albion College*
Pratibha Saxena, *University of Texas at Austin*
Jon Schnorr, *Pacific University*
Malcolm Schug, *University of North Carolina at Greensboro*
Deborah Schulman, *Lake Erie College*
Allan Showalter, *Ohio University*
Elaine Sia, *University of Rochester*
Robert Smith, *Nova Southeastern University*
Joyce Stamm, *University of Evansville*
Tara Stoulig, *Southeastern Louisiana University*
Julie Torruellas Garcia, *Nova Southeastern University*
Virginia Vandergon, *California State University*, Northridge
Charles Vigue, *University of New Haven*
Susan Walsh, *Rollins College*
Michael Watters, *Valparaiso University*
Roger Wartell, *Georgia Institute of Technology*

Matthew White, *Ohio University*
Dwayne Wise, *Mississippi State University*
Andrew Wood, *Southern Illinois University*
Mary Alice Yund, *UC Berkeley Extension*
Malcom Zellars, *Georgia State University*
Deborah Zies, *University of Mary Washington*

Tony Griffiths gostaria de agradecer às sugestões pedagógicas de David Suzuki, coautor das primeiras edições desta obra, o qual inspirou o público do mundo todo com seus ensinamentos na mídia. Agradece também a Jolie Mayer-Smith e Barbara Moon, que apresentaram a Tony o poder da abordagem construtivista aplicada ao ensino de genética. Sean Carroll agradece a Leanne Olds pela ajuda no trabalho de arte dos Capítulos 11, 12, 13, 14 e 20. John Doebley agradece aos seus colegas da University of Wisconsin Bill Engels, Carter Denniston e Jim Crow, que adaptaram a abordagem ao ensino da genética.

Os autores agradecem ainda à equipe da W. H. Freeman pelo árduo trabalho e pela paciência. Em particular, agradecemos à nossa editora de desenvolvimento e suplementos, Erica Champion; à editora de aquisição sênior Lauren Schultz; à editora de projetos sênior Jane O'Neill; e à copidesque Teresa Wilson. Agradecemos também a Susan Wein, supervisora de produção; Diana Blume, diretora de arte; Vicki Tomaselli, *designer* de texto e capa; Sheridan Sellers, paginador; Janice Donnola, coordenadora de ilustração; Jennifer MacMillan, direitos autorais; Amanda Dunning, editora executiva de mídia; e Alexandra Garrett, assistente editorial. Por fim, apreciamos especialmente os esforços de *marketing* e vendas de John Britch, diretor executivo de *marketing*, e de toda a equipe de vendas.

Sumário

1 A Revolução Genética

1.1 O surgimento da genética

Gregor Mendel | Um monge no jardim

Redescoberta de Mendel

Dogma central da biologia molecular

1.2 Após decifrar o código

Organismos-modelo

Ferramentas para a análise genética

1.3 Genética atual

Da genética clássica à genômica médica

Investigação da mutação e do risco de doença

Quando os pés de arroz ficam um pouco molhados

Evolução recente em humanos

2 Herança Monogênica

2.1 Padrões de herança monogênica

Experimentos pioneiros de Mendel

Lei de Mendel da segregação igual

2.2 Base cromossômica dos padrões de herança monogênica

Herança monogênica em diploides

Herança monogênica em haploides

2.3 Base molecular dos padrões de herança mendeliana

Diferenças estruturais entre os alelos no nível molecular

Aspectos moleculares da transmissão dos genes

Alelos no nível molecular

2.4 Alguns genes descobertos por meio da observação das proporções de

segregação
Um gene ativo no desenvolvimento da cor da flor
Um gene para o desenvolvimento das asas
Um gene para a ramificação das hifas
Previsão das proporções na progênie ou dos genótipos parentais por meio da aplicação dos princípios da herança monogênica

2.5 Padrões de herança monogênica ligada ao sexo
Cromossomos sexuais
Padrões de herança ligada ao sexo
Herança ligada ao X

2.6 Análise de heredogramas humanos
Distúrbios autossômicos recessivos
Distúrbios autossômicos dominantes
Polimorfismos autossômicos
Distúrbios recessivos ligados ao X
Distúrbios dominantes ligados ao X
Herança ligada ao Y
Cálculo de riscos na análise de heredogramas

3 Segregação Independente de Genes

3.1 Lei de Mendel de segregação independente

3.2 A segregação independente
Previsão das proporções da progênie
Utilização do teste do qui-quadrado em proporções mono-híbridas e di-híbridas
Síntese de linhagens puras
Vigor híbrido

3.3 Base cromossômica da segregação independente
Distribuição independente em organismos diploides
Segregação independente em organismos haploides
Segregação independente de combinações de genes autossômicos e ligados ao X
Recombinação

3.4 Herança poligênica

3.5 Genes de organelas | Herança independente do núcleo

Padrões de herança em organelas
Segregação citoplasmática
Mutações citoplasmáticas em humanos
mtDNA em estudos evolutivos

4 Mapeamento de Cromossomos Eucarióticos por Recombinação

4.1 Diagnóstico de ligação

Utilização da frequência de recombinantes para reconhecer ligação
Como os crossovers produzem recombinantes de genes ligados
Simbolismo e terminologia de ligação
Evidências de que o crossing over é um processo de quebra e reunião
Evidências de que o crossing over ocorre no estágio de quatro cromátides
Crossovers múltiplos podem incluir mais de duas cromátides

4.2 Mapeamento por frequência de recombinantes

Unidades de mapa
Cruzamento-teste de três pontos
Dedução da ordem dos genes por inspeção
Interferência
Utilização de proporções como diagnóstico

4.3 Mapeamento com marcadores moleculares

Polimorfismos de nucleotídio único
Polimorfismos de comprimento de sequência simples
Análise de recombinação com a utilização de marcadores moleculares

4.4 Mapeamento de centrômeros com tétrades lineares

4.5 Utilização do teste do qui-quadrado para inferir ligação

4.6 Cômputo de crossovers múltiplos não visualizados

Função de mapeamento
Fórmula de Perkins

4.7 Utilização de mapas com base em recombinação em conjunto com mapas físicos

4.8 Mecanismo molecular de crossover

5 A Genética das Bactérias e Seus Vírus

5.1 Trabalho com microrganismos

5.2 Conjugação bacteriana

Descoberta da conjugação

Descoberta do fator de fertilidade (F)

Linhagens Hfr

Mapeamento de cromossomos bacterianos

Plasmídios F carreadores de fragmentos genômicos

Plasmídios R

5.3 Transformação bacteriana

Natureza da transformação

Mapeamento cromossômico com a utilização de transformação

5.4 Genética de bacteriófagos

Infecção de bactérias por fagos

Mapeamento de cromossomos de fagos por meio da utilização de cruzamentos de fagos

5.5 Transdução

Descoberta da transdução

Transdução generalizada

Transdução especializada

Mecanismo de transdução especializada

5.6 Comparação de mapas físicos e mapas de ligação

6 Interação Gênica,

6.1 Interações de alelos de um único gene | Variações de dominância

Dominância completa e recessividade

Dominância incompleta
Codominância
Alelos letais recessivos

6.2 Interação dos genes nas vias
Vias bioquímicas de síntese em Neurospora
Interação gênica em outros tipos de vias bioquímicas

6.3 Inferência das interações gênicas
Segregação de mutantes com a utilização do teste de complementação
Análise de mutantes duplos de mutações aleatórias

6.4 Penetrância e expressividade

7 DNA | Estrutura e Replicação

7.1 DNA | O material genético
Descoberta da transformação
Experimento de Hershey-Chase

7.2 Estrutura do DNA
Estrutura do DNA antes de Watson e Crick
Dupla-hélice

7.3 Replicação semiconservativa
Experimento de Meselson-Stahl
Forquilha de replicação
DNA polimerases

7.4 Visão geral da replicação do DNA

7.5 Replissomo | Uma máquina de replicação notável
Deselicoidização da dupla-hélice
Montagem do replissomo | Início da replicação

7.6 Replicação em organismos eucariotos
Origens da replicação eucariótica
Replicação do DNA e ciclo celular de levedura

Origens de replicação em eucariotos superiores

7.7 Telômeros e telomerase | Término da replicação

8 RNA | Transcrição e Processamento

8.1 RNA

Experimentos iniciais sugerem um RNA intermediário

Propriedades do RNA

Classes de RNA

8.2 Transcrição

Visão geral | O DNA como modelo da transcrição

Estágios da transcrição

8.3 Transcrição em eucariotos

Iniciação da transcrição em eucariotos

Alongamento, término e processamento de pré-mRNA em eucariotos

8.4 Remoção de íntrons e recomposição de éxons

Pequenos RNA nucleares (snRNA) | O mecanismo de recomposição de éxons

Autorrecomposição de íntrons e o mundo do RNA

8.5 Pequenos RNA funcionais que regulam e protegem o genoma eucariótico

Os miRNA são importantes reguladores da expressão gênica

Os siRNA asseguram a estabilidade do genoma

Mecanismos semelhantes geram siRNA e miRNA

9 Proteínas e sua Síntese

9.1 Estrutura proteica

9.2 Código genético

Códigos sobrepostos versus não sobrepostos

Número de letras no códon

Utilização de supressores para demonstrar um código triplo

Degeneração do código genético

Como decifrar o código

Códons de parada

9.3 tRNA | O adaptador

Tradução do códon pelo tRNA

Novamente a degeneração

9.4 Ribossomos

Características do ribossomo

Iniciação, alongamento e término da tradução

Mutações supressoras sem sentido

9.5 Proteoma

A recomposição alternativa gera isoformas de proteínas

Eventos pós-tradução

10 Isolamento e Manipulação de Genes

10.1 Visão geral | Isolamento e amplificação de fragmentos específicos de DNA

10.2 Produção de moléculas de DNA recombinante

O DNA genômico pode ser cortado antes da clonagem

A reação da cadeia de polimerase amplifica regiões selecionadas do DNA in vitro

Cópias do DNA podem ser sintetizadas a partir de mRNA

Ligação do DNA doador e do vetor

Amplificação do DNA doador em uma célula bacteriana

Construção de bibliotecas genômicas e de cDNA

10.3 Utilização de sondas moleculares para encontrar e analisar um clone de interesse específico

Detecção de clones específicos com a utilização de sondas

Detecção de clones específicos por complementação funcional

Análise de DNA por Southern e Northern blot

10.4 Determinação da sequência de bases de um segmento de DNA

10.5 Alinhamento genético e mapas físicos para isolar genes específicos

Utilização de clonagem posicional para identificar o gene de uma doença humana

Utilização do mapeamento fino para identificar genes

10.6 Engenharia genética

Engenharia genética em Saccharomyces cerevisiae

Engenharia genética em plantas

Engenharia genética em animais

11 Regulação da Expressão Gênica em Bactérias e seus Vírus

11.1 Regulação gênica

Base da regulação da transcrição procariótica | Interruptores genéticos

Primeira observação do circuito regulador lac

11.2 Descoberta do sistema lac | Controle negativo

Genes controlados juntos

Evidências genéticas do operador e do repressor

Evidências genéticas em relação à alosteria

Análise genética do promotor lac

Caracterização molecular do repressor Lac e do operador *lac*

11.3 Repressão catabólica do óperon lac | Controle positivo

Base da repressão catabólica de lac | Escolha do melhor açúcar a ser metabolizado

Estrutura dos sítios-alvo no DNA

Resumo do óperon lac

11.4 Duplo controle positivo e negativo | Óperon da arabinose

11.5 Vias metabólicas e níveis adicionais de regulação | Atenuação

11.6 Ciclos de vida de bacteriófagos | Mais reguladores, óperons complexos

Anatomia molecular do interruptor genético

Ligação sequência-específica de proteínas reguladoras ao DNA

11.7 Fatores sigma alternativos regulam grandes conjuntos de genes

12 Regulação da Expressão Gênica em Eucariotos

12.1 Regulação da transcrição em eucariotos | Visão geral

12.2 Lições das leveduras | Sistema GAL

Gal4 regula múltiplos genes por meio de sequências de ativação upstream
Proteína Gal4 apresenta domínios de ligação ao DNA e ativação separados
Atividade da Gal4 é regulada fisiologicamente
Funções da Gal4 na maior parte dos eucariotos
Ativadores recrutam o maquinário de transcrição
Controle do tipo reprodutivo de leveduras | Interações combinatórias

12.3 Cromatina dinâmica
Proteínas remodeladoras da cromatina e ativação gênica
Modificação das histonas
Metilação das histonas pode ativar ou reprimir a expressão gênica
Herança das modificações de histonas e estrutura da cromatina
Variantes de histonas
Metilação do DNA | Outra marca hereditária que influencia a estrutura da cromatina

12.4 Ativação de genes em um ambiente de cromatina
Acentuassomo de β-interferona
Isoladores de bloqueio de acentuador

12.5 Inativação a longo prazo de genes em um ambiente de cromatina
Alteração do tipo reprodutivo e silenciamento gênico
Comparação da heterocromatina e da eucromatina
Variegação por efeito de posição em Drosophila revela vizinhanças genômicas
Análise genética de PEV revela proteínas necessárias para a formação de heterocromatina

12.6 Silenciamento gênero-específico de genes e cromossomos inteiros
Imprinting genômico explica alguns padrões de herança incomuns
Mas o que dizer sobre Dolly e outros mamíferos clonados?
Silenciamento de um cromossomo inteiro | Inativação do cromossomo X

12.7 Repressão gênica pós-transcricional pelos miRNA

13 Controle Genético do Desenvolvimento

13.1 Abordagem genética do desenvolvimento
Análise mutacional do desenvolvimento inicial de Drosophila
Utilização do conhecimento de um organismo-modelo para acelerar a descoberta de genes de desenvolvimento em outros organismos

13.2 Toolkit genética para o desenvolvimento de Drosophila
Classificação dos genes pela função no desenvolvimento
Genes homeóticos e identidade segmentar
Organização e expressão dos genes Hox
Homeoboxe
Grupos de genes Hox controlam o desenvolvimento na maior parte dos animais
13.3 Definição da toolkit inteira
Eixos anteroposterior e dorsoventral
Expressão dos genes toolkit
13.4 Regulação espacial da expressão gênica no desenvolvimento
Gradientes maternos e ativação gênica
Desenho das listras | Integração dos estímulos das proteínas gap
Diferenciação dos segmentos | Integração dos estímulos de Hox
13.5 Regulação pós-transcricional da expressão gênica no desenvolvimento
Recomposição do RNA e determinação do sexo em Drosophila
Regulação da tradução do mRNA e linhagem celular em C. elegans
Controle da tradução no embrião inicial
O nematódeo Caenorhabditis elegans como um modelo para decisões do destino de linhagens celulares
Controle da cronologia do desenvolvimento por miRNA em C. elegans e outras espécies
13.6 De moscas a dedos, penas e placas do assoalho | Os muitos papéis dos genes toolkit
13.7 Desenvolvimento e doença
Polidactilia
Holoprosencefalia
Câncer como uma doença do desenvolvimento
14 Genomas e Genômica
14.1 Revolução genômica
14.2 Obtenção da sequência de um genoma
Transformação das leituras de sequências em uma sequência montada

Sequenciamento do genoma inteiro
WGS tradicional
Sequenciamento shotgun de última geração do genoma inteiro
Montagem da sequência do genoma inteiro

14.3 Bioinformática | Significado da sequência genômica
Natureza do conteúdo da informação do DNA
Dedução dos genes codificadores de proteínas a partir da sequência genômica

14.4 Estrutura do genoma humano
Elementos funcionais não codificadores no genoma

14.5 Genômica comparativa de seres humanos com outras espécies
Inferência filogenética
Camundongos e humanos
Genômica comparativa de chimpanzés e seres humanos

14.6 Genômica comparativa e medicina de seres humanos
Exoma e genômica personalizada
Genômica comparativa de E. coli não patogênica e patogênica

14.7 Genômica funcional e genética reversa
"Omas"
Genética reversa

15 Genoma Dinâmico | Elementos de Transposição

15.1 Descoberta dos elementos de transposição no milho
Experimentos de McClintock | O elemento *Ds*
Elementos autônomos e não autônomos
Elementos de transposição | Apenas no milho?

15.2 Elementos de transposição em procariotos
Sequências de inserção bacteriana
Transpósons procarióticos
Mecanismo de transposição

15.3 Elementos de transposição em eucariotos

Classe 1 | Retrotranspósons
Classe 2 | Transpósons de DNA
Utilidade dos transpósons de DNA para a descoberta de genes

15.4 O genoma dinâmico | Mais elementos de transposição do que jamais se imaginou
Grandes genomas são predominantemente elementos de transposição
Elementos de transposição no genoma humano
As gramíneas | Retrotranspósons LTR desenvolvem-se em grandes genomas
Abrigos seguros

15.5 Regulação epigenética de elementos de transposição pelo hospedeiro
Vigilância do genoma em animais e bactérias

16 Mutação, Reparo e Recombinação

16.1 Consequências fenotípicas das mutações no DNA
Tipos de mutação de ponto
Consequências moleculares das mutações de ponto em uma região codificadora
Consequências moleculares das mutações de ponto em uma região não codificadora

16.2 Base molecular das mutações espontâneas
Teste de flutuação de Luria e Delbrück
Mecanismos de mutações espontâneas
Mutações espontâneas em seres humanos | Doenças por repetições de trinucleotídios

16.3 Base molecular das mutações induzidas
Mecanismos de mutagênese
Teste de Ames | Avaliação de mutágenos em nosso ambiente

16.4 Mecanismos biológicos de reparo
Reversão direta de DNA danificado
Reparo por excisão de base
Reparo por excisão de nucleotídio
Reparo pós-replicação | Reparo de malpareamento
Reparo propenso a erro | Síntese de DNA translesão
Reparo de quebras bifilamentares

Envolvimento do reparo DSB na recombinação meiótica

16.5 Câncer | Uma importante consequência fenotípica da mutação

Como as células cancerosas diferem das células normais

Mutações em células cancerosas

17 Alterações Cromossômicas em Grande Escala

17.1 Alterações no número de cromossomos

Euploidia aberrante

Aneuploidia

Conceito de balanço gênico

17.2 Alterações na estrutura dos cromossomos

Deleções

Duplicações

Inversões

Translocações recíprocas

Translocações robertsonianas

Aplicações de inversões e translocações

Rearranjos e câncer

Identificação de mutações cromossômicas pela genômica

17.3 Incidência geral de mutações cromossômicas humanas

18 Genética de Populações

18.1 Detecção da variação genética

Polimorfismos de nucleotídio único

Microssatélites

Haplótipos

Outras fontes e outros tipos de variação

Projeto HapMap

18.2 Conceito de pool gênico e lei de Hardy-Weinberg

18.3 Sistemas de acasalamento

Cruzamento preferencial
Isolamento pela distância
Endocruzamento
Coeficiente de endocruzamento
Tamanho da população e endocruzamento

18.4 Variação genética e sua medida

18.5 Modulação da variação genética

Novos alelos na população | Mutação e migração
Recombinação e desequilíbrio de ligação
Deriva genética e tamanho da população
Seleção
Tipos de seleção
Equilíbrio entre mutação e deriva
Equilíbrio entre mutação e seleção

18.6 Aplicações biológicas e sociais

Genética da conservação
Cálculo dos riscos de doenças
DNA forense
Procura no Google por seus parceiros de DNA

19 Herança de Traços Complexos

19.1 Medida da variação quantitativa

Tipos de traço e herança
Média
Variância
Distribuição normal

19.2 Modelo genético simples para os traços quantitativos

Desvios genéticos e ambientais
Variâncias genética e ambiental
Correlação entre variáveis

19.3 Herdabilidade no sentido amplo | Natureza versus criação

Medida da herdabilidade em seres humanos a partir de estudos com gêmeos

19.4 Herdabilidade no sentido restrito | Previsão dos fenótipos
Ação gênica e transmissão de variação genética
Efeitos aditivos e da dominância
Modelo com aditividade e dominância
Herdabilidade no sentido restrito
Previsão dos fenótipos da descendência
Seleção de traços complexos

19.5 Mapeamento de QTL em populações com heredogramas conhecidos
Método básico
Do QTL ao gene

19.6 Mapeamento de associação em populações de cruzamento aleatório
Método básico
GWA, genes, doença e herdabilidade

20 Evolução de Genes e Traços

20.1 Evolução por seleção natural

20.2 Seleção natural em ação | Caso exemplar
Vantagem seletiva de Hb^S
Origens moleculares de Hb^S

20.3 Evolução molecular | Teoria neutra
Desenvolvimento da teoria neutra
Taxa de substituições neutras
Assinatura da seleção purificadora no DNA

20.4 Seleção cumulativa e vias de múltiplas etapas até a alteração funcional
Vias de múltiplas etapas na evolução
Assinatura de seleção positiva nas sequências de DNA

20.5 Evolução morfológica
Alterações adaptativas em uma proteína reguladora de pigmento
Inativação gênica
Evolução de sequência reguladora

Perda de características por meio da evolução de sequências reguladoras
Evolução reguladora em seres humanos

20.6 Origem de novos genes e funções proteicas

Expansão do número de genes
Destino de genes duplicados

Guia Resumido de Organismos-modelo

Apêndice A: Nomenclatura genética

Apêndice B: Recursos de bioinformática para genética e genômica

Glossário

Respostas dos Problemas Selecionados

INTRODUÇÃO À
GENÉTICA

CAPÍTULO 1

A Revolução Genética

O DNA (ácido desoxirribonucleico) é a molécula que codifica a informação genética. Os filamentos com 4 diferentes bases químicas no DNA armazenam a informação genética de modo muito semelhante ao modo como as séries de 0 e 1 armazenam as informações no código de computador. (Sergey Nivens/Shutterstock.)

TÓPICOS

1.1 Surgimento da genética
1.2 Após decifrar o código
1.3 Genética atual

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Descrever o modo como a genética moderna se desenvolveu
- Listar os principais constituintes celulares envolvidos na expressão e na ação dos genes
- Fornecer alguns exemplos de como a genética influenciou a medicina moderna, a agricultura e a evolução.

A genética é um tipo de ciência da informação. Os geneticistas tentam compreender as regras que controlam a transmissão da informação genética em 3 níveis: do genitor à descendência, dentro das famílias; do DNA à ação dos genes dentro das células e entre elas; e ao longo de muitas gerações dentro de populações de organismos. Esses três focos da genética são conhecidos como genética da transmissão, genética molecular do desenvolvimento e genética evolutiva de populações. As três partes deste texto examinam esses três focos da genética.

A ciência da genética surgiu há pouco mais de 100 anos. Desde aquela época, a genética alterou profundamente nossa compreensão sobre a vida, desde o nível da célula individual até aquele de uma população de organismos que se desenvolve ao longo de milhões de anos. Em 1900, William Bateson, um proeminente biólogo britânico, escreveu de modo profético que a "determinação exata das leis da hereditariedade provavelmente causará mais alterações na percepção do ser humano sobre o mundo e em seu poder sobre a natureza do que qualquer outro avanço no conhecimento natural que possa ser previsto". Em todo este texto, você observará a realização da previsão de Bateson. A genética desencadeou uma revolução nas ciências biológicas e na sociedade em geral.

Neste primeiro capítulo, revisaremos brevemente a história da genética e, ao fazê-lo, revisaremos alguns dos conceitos básicos dessa disciplina, os quais foram descobertos ao longo dos últimos 100 anos. Depois disso, observaremos alguns exemplos de como a análise genética está sendo aplicada atualmente em

problemas críticos na biologia, na agricultura e na saúde humana. Você verá como as pesquisas contemporâneas em genética integram conceitos descobertos há décadas com recentes avanços tecnológicos. Verá também que atualmente a genética é um campo de investigação dinâmico, no qual novas descobertas estão desenvolvendo continuamente a nossa compreensão sobre o mundo biológico.

1.1 O surgimento da genética

Ao longo de toda a história registrada, as pessoas ao redor do mundo compreendiam que “semelhantes geram semelhantes”. Os filhos se assemelham aos seus genitores, a semente de uma árvore que contém frutos saborosos por sua vez crescerá como uma árvore carregada com frutos saborosos e mesmo membros de grupos de lobos demonstram semelhanças familiares (Figura 1.1). Embora as pessoas confiassem nessas observações, elas pensavam a respeito do mecanismo subjacente. A tribo ameríndia Hopi, do sudoeste dos EUA, compreendia que, se eles plantassem um grão de milho vermelho em seus campos, ele cresceria com uma planta que também forneceria grãos vermelhos. O mesmo era verdadeiro em relação aos grãos azuis, brancos ou amarelos. Assim, eles consideravam o grão como uma mensagem aos deuses na Terra a respeito do tipo de milho que os fazendeiros Hopi esperavam colher. Ao receber essa mensagem, os deuses fielmente devolveriam a eles uma planta que produziria grãos da cor desejada.

FIGURA 1.1 Os grupos familiares nos lobos-cinzentos demonstram semelhanças familiares em relação às cores e ao padrão do pelo (*Parte superior*, altrendo nature/Getty Images; *parte inferior*, Bev McConnell/Getty Images.)

Nos anos 1800, na Europa, horticultores, criadores de animais e biólogos também tentavam explicar as semelhanças entre os genitores e a descendência. Naquela época, uma visão comum era a **teoria da mistura** da herança ou a crença de que a herança atuava como uma mistura de líquidos, semelhante ao que ocorre com as tintas: tintas vermelha e branca, quando misturadas, originam cor-de-rosa. Assim, seria esperado que o filho de um genitor alto com um genitor baixo

crescesse até uma altura intermediária. Embora a teoria da mistura aparentasse atuar algumas vezes, também estava claro que havia exceções, como crianças altas nascidas de genitores de estatura mediana. A teoria da mistura também não fornecia nenhuma explicação por meio da qual os "líquidos da hereditariedade", após a mistura, pudessem ser separados — as tintas vermelha e branca não podem ser reconstituídas a partir do cor-de-rosa. Portanto, a expectativa a longo prazo da teoria da mistura ao longo de muitas gerações de intercruzamento entre indivíduos era que todos os membros da população irão expressar o mesmo valor médio de um traço. Claramente, não é assim que a natureza atua. As populações humanas apresentam pessoas com uma diversidade de estaturas, de baixas a altas, e não apresentamos uma estatura média única, apesar das muitas gerações das populações humanas que residiram sobre a Terra.

Gregor Mendel | Um monge no jardim

Enquanto os méritos e as falhas da teoria da mistura estavam sendo debatidos, Gregor Mendel, um monge austríaco, estava trabalhando para compreender as regras que controlam a transmissão dos traços dos genitores para a descendência após a hibridização entre diferentes variedades de ervilha (Figura 1.2). O ambiente do seu trabalho era o jardim do monastério, na cidade de Brünn, Áustria (atualmente Brno, na República Tcheca). De 1856 a 1863, Mendel realizou a polinização cruzada ou o intercruzamento de diferentes variedades de ervilha. Um de seus experimentos envolveu o cruzamento entre uma variedade com flores roxas e outra com flores brancas (Figura 1.3). Mendel registrou que toda a primeira geração híbrida da descendência desse cruzamento apresentava flores roxas, bem como um dos genitores. Não havia mistura. Em seguida, Mendel autopolinizou as plantas híbridas de primeira geração e cultivou uma segunda geração da descendência. Entre a progênie, ele observou plantas com flores roxas, bem como plantas com flores brancas. Das 929 plantas, ele registrou 705 com flores roxas e 224 com flores brancas (Figura 1.4). Ele observou que havia aproximadamente três plantas com flores roxas para cada planta com flores brancas.

Como Mendel explicou os seus resultados? Claramente, a teoria da mistura não

funcionaria, tendo em vista que aquela teoria prevê um grupo uniforme de plantas híbridas da primeira geração com flores roxo-claras. Portanto, Mendel propôs que os fatores que controlam os traços atuam como *partículas*, em vez de líquidos, e que tais partículas não se misturam em conjunto, mas são transmitidas intactas de uma geração para a próxima. Atualmente, as partículas de Mendel são conhecidas como **genes**.

FIGURA 1.2 Gregor Mendel era um monge austríaco que descobriu as leis da herança. (James King-Holmes/Science Source.)

Um dos experimentos de Mendel

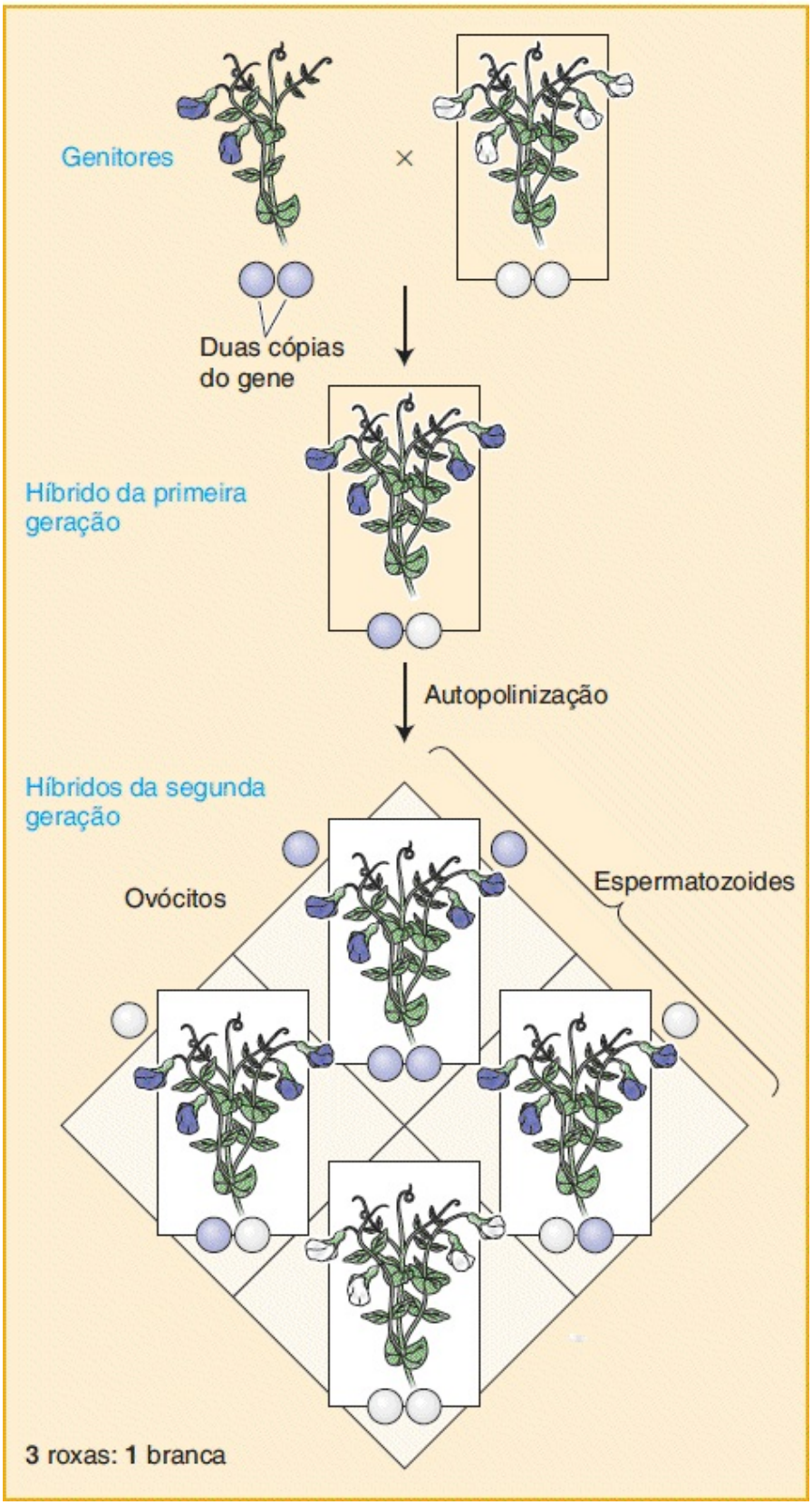


FIGURA 1.3 O esquema de cruzamento do experimento de Mendel, envolvendo o cruzamento de variedades de ervilhas com flores roxas e brancas. Os *círculos roxos* e *brancos* significam as variantes gênicas para a cor da flor roxa *versus* branca. Os gametas carreiam uma cópia do gene. Cada uma das plantas carreia duas cópias do gene. O × significa uma polinização cruzada entre as plantas com flores roxas e brancas.

Mendel propôs que cada planta da ervilha apresenta duas cópias do gene que codifica a cor da flor em cada uma das células do corpo da planta (**células somáticas**). Entretanto, quando a planta forma as células sexuais, ou **gametas** (ovócitos e espermatozoides), apenas uma cópia do gene entra nessas células reprodutivas (ver Figura 1.3). Em seguida, quando o ovócito e o espermatozoide se unem para dar início a um novo ser, mais uma vez haverá duas cópias do gene da cor da flor em cada célula do corpo da planta.

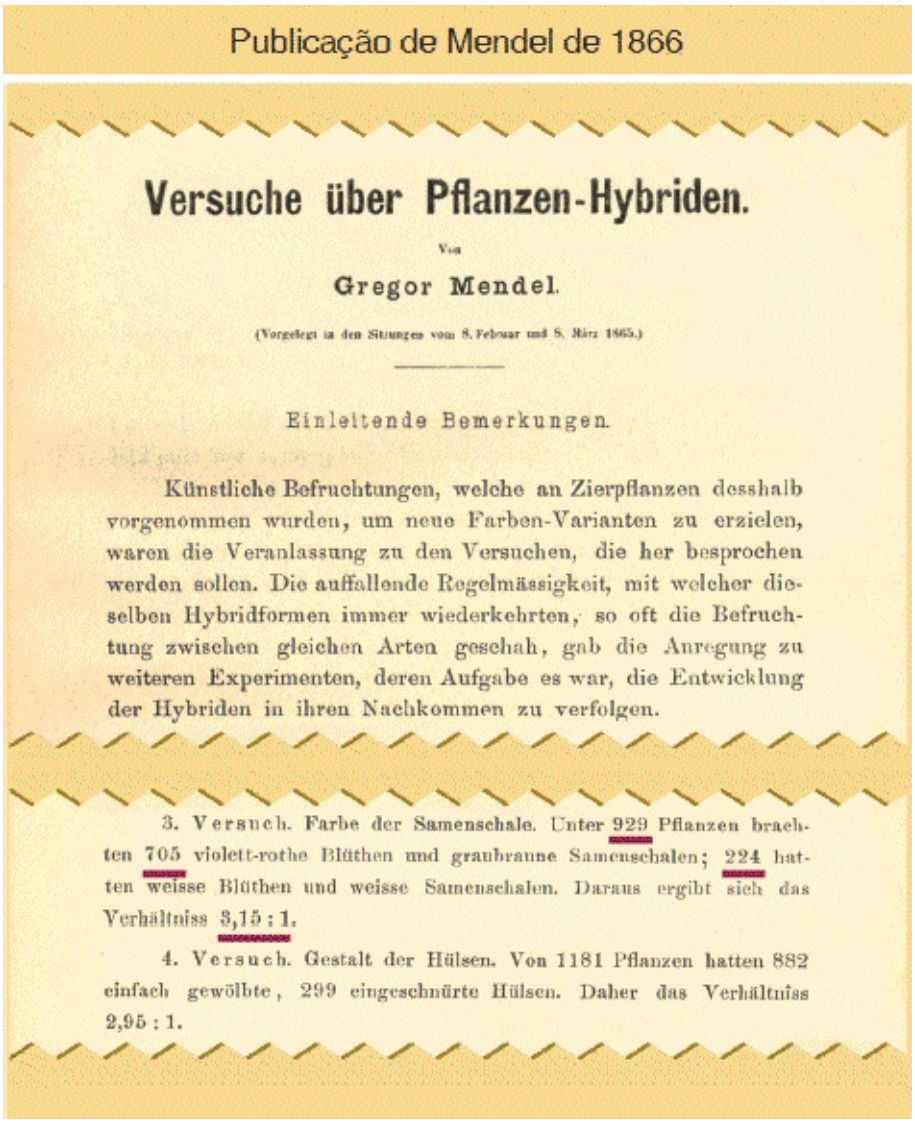


FIGURA 1.4 Trechos da publicação de Mendel de 1866, *Versuche über Pflanzen-Hybriden* (Experimentos com hibridização de plantas). (Abadia Agostiniana em Old Brno. Cortesia de Masaryk University, Mendel Museum.)

Mendel apresentou algumas percepções adicionais. Ele propôs que o gene para a cor da flor ocorre em duas variantes genéticas, ou **alelos**: um que condiciona as flores roxas e outro que condiciona as brancas. Propôs também que o alelo roxo do gene da cor da flor é **dominante** em relação ao alelo branco, de modo que uma planta com um alelo roxo e um alelo branco apresentaria flores roxas. Apenas as plantas com dois alelos brancos apresentariam flores brancas (ver Figura 1.3). As duas conclusões de Mendel, (1) de que os genes se comportavam como partículas

que não se misturam em conjunto e (2) de que um alelo é dominante em relação ao outro, possibilitaram que ele explicasse a ausência de mistura nos híbridos da primeira geração e o reaparecimento das plantas com flores brancas nos híbridos da segunda geração, com uma proporção de 3:1 de plantas com flores roxas e brancas. Esse avanço revolucionário na nossa compreensão sobre a herança será totalmente discutido no Capítulo 2.

Como Mendel acertou quando tantos outros antes dele estavam errados? Mendel escolheu um bom organismo e bons traços para estudar. Todos os traços que ele estudou eram controlados por genes únicos. Os traços que são controlados por diversos genes, como ocorre com muitos traços, não teriam possibilitado que ele descobrisse as leis da herança tão facilmente. Mendel também era um observador cuidadoso e manteve registros detalhados de cada um dos seus experimentos. Por fim, Mendel era um pensador criativo, capaz de raciocinar bem além das ideias de sua época.

A teoria da herança particulada de Mendel foi publicada em 1866 nos *Proceedings of the Natural History Society de Brünn* (ver Figura 1.4). Naquela época, seu trabalho chamou a atenção e foi lido por alguns biólogos, mas as suas implicações e a sua importância foram subestimadas durante mais de 30 anos. Ao contrário de Charles Darwin, cuja teoria da evolução por meio da seleção natural o tornou mundialmente reconhecido da noite para o dia, quando Mendel morreu, em 1884, era relativamente desconhecido no mundo da ciência. Conforme o bioquímico Erwin Chargaff mencionou: “Existem pessoas que parecem ter nascido com um manto de invisibilidade. Mendel era uma delas.”

> CONCEITO-CHAVE Gregor Mendel demonstrou que os genes se comportavam como partículas, não como líquidos.

Redescoberta de Mendel

Conforme diz a lenda, quando o biólogo britânico William Bateson (Figura 1.5) embarcou em um trem dirigindo-se a uma conferência em Londres, em 1900, ele

não tinha ideia do quão profundamente o seu mundo iria mudar durante a breve jornada. Bateson carregava consigo uma cópia do artigo de Mendel de 1866 sobre a hibridização das variedades de plantas. O estudioso soubera recentemente que biólogos na Alemanha, na Holanda e na Áustria haviam reproduzido, cada um de modo independente, a proporção de 3:1 de Mendel e que cada um citava o trabalho original de Mendel. Esse trio havia redescoberto as leis de Mendel acerca da herança. Bateson precisava ler o artigo de Mendel. Quando desceu do trem, o biólogo tinha uma nova missão na vida. Havia compreendido que o mistério da herança fora solucionado. Logo, tornou-se um apóstolo persistente das leis de Mendel sobre a herança. Alguns anos depois, em 1905, Bateson cunhou o termo **genética**: o estudo da herança. A revolução genética havia começado.

FIGURA 1.5 William Bateson, o zoólogo e evolucionista britânico que introduziu o termo *genética* para o estudo da herança e que promoveu o trabalho de Mendel. (SPL/Science Source.)

Quando as leis de Mendel sobre a herança foram redescobertas, em 1900, foi liberado um grande fluxo de novos pensamentos e novas ideias. O mendelismo tornou-se o princípio de organização para uma grande parte da biologia. Havia muitas novas questões a ser formuladas a respeito da herança. A Tabela 1.1 resume a cronologia das descobertas influentes realizadas ao longo das décadas subsequentes e os capítulos deste livro que abrangem cada um desses tópicos.

Vejamos brevemente algumas questões que transformaram as ciências biológicas e suas respostas.

Tabela 1.1 Eventos-chave na história da genética.

Ano	Evento	Capítulos
1865	Gregor Mendel demonstrou que os traços são controlados por fatores distintos, atualmente conhecidos como genes	2, 3
1869	Friedrich Miescher isolou o DNA dos núcleos de leucócitos	7
1903	Walter Sutton e Theodor Boveri formularam a hipótese de que os cromossomos são os elementos hereditários	4
1905	William Bateson introduziu o termo *genética* para descrever o estudo da herança	2
1908	G. H. Hardy e Wilhelm Weinberg propuseram a lei de Hardy-Weinberg, o fundamento da genética de populações	18
1910	Thomas H. Morgan demonstrou que os genes estão localizados nos cromossomos	4
1913	Alfred Sturtevant criou um mapa da ligação genética do cromossomo X da *Drosophila*, o primeiro mapa genético	4

1918	Ronald Fisher propôs que fatores mendelianos múltiplos podem explicar a variação contínua em relação aos traços, fundando o campo da genética quantitativa	19
1931	Harriet Creighton e Barbara McClintock demonstraram que o *crossing over* é a causa da recombinação	4, 16
1941	Edward Tatum e George Beadle propuseram a hipótese de um gene—um polipeptídio	6
1944	Oswald Avery, Colin MacLeod e Maclyn McCarty forneceram evidências persuasivas de que o DNA é o material genético nas células bacterianas	7
1946	Joshua Lederberg e Edward Tatum descobriram a conjugação bacteriana	5
1948	Barbara McClintock descobriu elementos móveis (transpósons) que se movimentam de um local para outro no genoma	15
1950	Erwin Chargaff demonstrou que a composição do DNA segue algumas regras simples em relação às quantidades relativas de A, C, G e T	7
1952	Alfred Hershey e Martha Chase comprovaram que o DNA é a molécula que codifica as informações genéticas	7

1953	James Watson e Francis Crick determinaram que o DNA forma uma dupla-hélice	7
1958	Matthew Meselson e Franklin Stahl demonstraram a natureza semiconservativa da replicação do DNA	7
1958	Jérôme Lejeune descobriu que a síndrome de Down resultava de uma cópia extra do cromossomo 21	17
1961	François Jacob e Jacques Monod propuseram que os níveis enzimáticos nas células são controlados por mecanismos de retroalimentação	11
1961 a 1967	Marshall Nirenberg, Har Gobind Khorana, Sydney Brenner e Francis Crick “desvendaram” o código genético	9
1968	Motoo Kimura propôs a teoria neutra da evolução molecular	18, 20
1977	Fred Sanger, Walter Gilbert e Allan Maxam inventaram métodos para a determinação das sequências de nucleotídios das moléculas de DNA	10
1980	Christiane Nüsslein-Volhard e Eric F. Wieschaus definiram o complexo de genes que regulam o desenvolvimento do plano corporal na *Drosophila*	13

1989	Francis Collins e Lap-Chee Tsui descobriram o gene que causa a fibrose cística	4, 10
1993	Victor Ambrose e colegas descobriram o primeiro microRNA	13
1995	Publicada a primeira sequência de genoma de um organismo vivo (*Haemophilus influenzae*)	14
1996	Publicada a primeira sequência de genoma de um eucarioto (*Saccharomyces cerevisiae*)	14
1998	Publicada a primeira sequência de genoma de um animal (*Caenorhabditis elegans*)	14
2000	Publicada a primeira sequência de genoma de uma planta (*Arabidopsis thaliana*)	14
2001	Publicada pela primeira vez a sequência do genoma humano	14
2006	Andrew Fire e Craig Mello vencem o prêmio Nobel por sua descoberta do gene que é silenciado pelo RNA de filamento duplo	8
2012	John Gurdon e Shinya Yamanaka vencem o prêmio Nobel por sua descoberta de que apenas 4 genes reguladores conseguem converter células adultas em	8, 12

células-tronco

Onde na célula estão os genes de Mendel? A resposta veio em 1910, quando Thomas H. Morgan, da Columbia University, em Nova York, demonstrou que os genes de Mendel estão localizados nos cromossomos: ele comprovou a **teoria cromossômica** da herança. A ideia não era novidade. Walter Sutton, que cresceu em uma fazenda no Kansas e posteriormente atuou como cirurgião para o exército norte-americano durante a Primeira Guerra Mundial, havia proposto a teoria cromossômica da herança em 1903. Theodor Boveri, um biólogo alemão, propôs tal teoria independentemente na mesma época. Era uma hipótese persuasiva, mas não havia dados experimentais para ampará-la. Isso mudou em 1910, quando Morgan comprovou a teoria cromossômica da herança com a utilização da genética mendeliana e a mosca-das-frutas como seu organismo experimental. No Capítulo 4, você retraçará os experimentos de Morgan que comprovaram que os genes estão nos cromossomos.

Os genes mendelianos conseguem explicar a herança de traços continuamente variáveis, como a estatura humana? Enquanto as proporções de segregação de 3:1 podem ser diretamente observadas em relação a traços simples, como a cor da flor, muitos traços demonstram uma variação contínua de valores em híbridos da segunda geração, sem proporções simples como 3:1. Em 1918, Ronald Fisher, o estatístico e geneticista britânico, resolveu como os genes mendelianos explicavam a herança de traços continuamente variáveis, como a estatura nas pessoas (Figura 1.6). A ideia central de Fisher era de que os traços contínuos são, cada um, controlados por genes mendelianos múltiplos. A percepção de Fisher é conhecida como **hipótese multifatorial**. No Capítulo 19, dissecaremos o modelo matemático e a evidência experimental em relação à hipótese de Fisher.

Como os genes atuam dentro das células de modo que lhes é possibilitado o controle de diferentes estados em relação a um traço como a cor das flores? Em 1941, Edward Tatum e George Beadle propuseram que os genes codificam

enzimas. Com a utilização do fungo do pão (*Neurospora crassa*) como seu organismo experimental, eles demonstraram que os genes codificam as enzimas que realizam funções metabólicas dentro das células (Figura 1.7). No caso da ervilha, existe um gene que codifica uma enzima necessária para produzir o pigmento roxo nas células de uma flor. A descoberta de Tatum e Beadle se tornou conhecida como a **hipótese de um gene–uma enzima**. Você verá como eles desenvolveram tal hipótese no Capítulo 6.

FIGURA 1.6 Estudantes no Connecticut Agricultural College em 1914 demonstram uma diversidade de estaturas. Ronald Fisher propôs que traços continuamente variáveis, como a estatura humana, são controlados por genes mendelianos múltiplos. (A. F. Blakeslee, "Corn and Men," *Journal of Heredity* 5, 11, 1914, 511-518.)

Qual é a natureza física do gene? Os genes são compostos por proteínas, ácido nucleico ou outra substância? Em 1944, Oswald Avery, Colin MacLeod e Maclyn McCarty ofereceram a primeira evidência experimental persuasiva de que os genes são feitos de ácido desoxirribonucleico (DNA). Eles demonstraram que o DNA extraído de uma cepa virulenta de bactérias carreava as informações genéticas necessárias para transformar uma cepa não virulenta em uma cepa virulenta. Você aprenderá no Capítulo 7 como exatamente eles demonstraram isso.

Como as moléculas de DNA conseguem armazenar informação? Na década de 1950, havia um certo tipo de corrida entre diversos grupos de geneticistas e

químicos para responder a essa questão. Em 1953, James Watson e Francis Crick, que trabalhavam na Cambridge University, na Inglaterra, venceram a corrida. Eles determinaram que a estrutura molecular do DNA se encontrava na forma de uma dupla-hélice: dois filamentos de DNA enrolados lado a lado em uma espiral. Sua estrutura da dupla-hélice é semelhante a uma escada torcida (Figura 1.8). As laterais da escada são compostas por grupos de açúcar e fosfato. Os degraus, por 4 bases: **adenina (A)**, **timina (T)**, **guanina (G)** e **citosina (C)**. As bases estão voltadas para o centro e cada base está ligada por uma ponte de hidrogênio com a base voltada para ela no filamento oposto. A adenina em um filamento sempre está pareada com a timina no outro, por meio de uma *dupla ligação de hidrogênio*, enquanto a guanina sempre está pareada com a citosina por meio de uma *tripla ligação de hidrogênio*. A especificidade da ligação tem por base os formatos e as cargas elétricas **complementares** das bases. A sequência de A, T, G e C representa a informação codificada pela molécula de DNA. Você aprenderá no Capítulo 7 como tudo isso ocorreu.

Como os genes são regulados? As células necessitam de mecanismos para ligar ou desligar os genes em tipos de células e tecidos específicos e em ocasiões específicas durante o desenvolvimento. Em 1961, François Jacob e Jacques Monod realizaram uma descoberta conceitual sobre essa questão. Trabalhando com os genes necessários para metabolizar o açúcar lactose na bactéria *Escherichia coli*, eles demonstraram que os genes apresentam **elementos regulatórios** que comandam a **expressão gênica**, ou seja, se um gene está ligado ou desligado (Figura 1.9). Os elementos regulatórios são sequências de DNA específicas às quais uma proteína regulatória se liga e atua como um ativador ou um repressor da expressão do gene. No Capítulo 11, você explorará a lógica por trás dos experimentos de Jacob e Monod com *E. coli*, e, no 12, você explorará os detalhes da regulação gênica em eucariotos.

Como as informações armazenadas no DNA são decodificadas para sintetizar proteínas? Embora a descoberta da estrutura helicoidal dupla do DNA tenha sido um divisor de águas para a biologia, muitos detalhes ainda eram desconhecidos. Precisamente como as informações eram codificadas no DNA e como elas eram decodificadas para formar as enzimas que Tatum e Beadle

haviam demonstrado ser desencadeadoras da ação gênica permanecia desconhecido. Entre os anos de 1961 e 1967, equipes de geneticistas moleculares e químicos que trabalhavam em diversos países responderam a essas questões quando “desvendaram o código genético”. Isso significa que eles deduziram como uma sequência de nucleotídios no DNA, cada qual com uma de quatro bases diferentes (A, T, C ou G), codifica o conjunto de 20 aminoácidos diferentes que formam as proteínas. Eles também descobriram que existe uma molécula mensageira, feita de ácido ribonucleico (RNA), que carreia as informações do DNA no núcleo para o citoplasma, no qual as proteínas são sintetizadas. Em 1967, o fluxograma básico em relação à transmissão da informação nas células foi conhecido. Esse fluxograma é considerado o dogma central da biologia molecular.

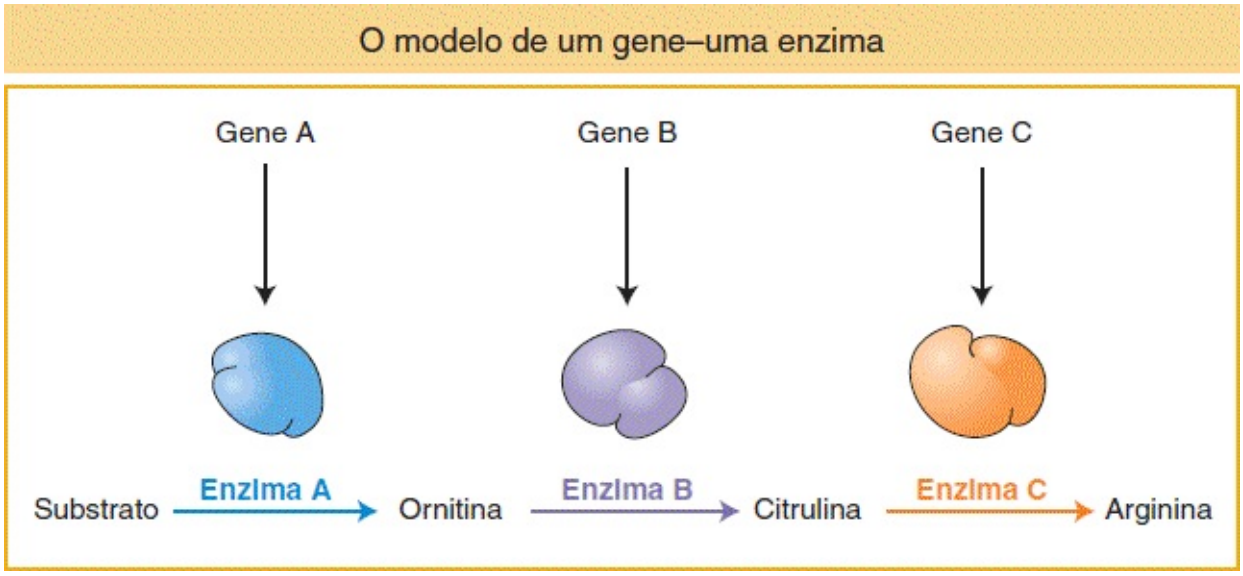


FIGURA 1.7 O modelo de um gene–uma enzima propôs que os genes codificam enzimas que realizam funções bioquímicas dentro das células. Tatum e Beadle propuseram esse modelo com base no estudo da síntese da arginina (um aminoácido) no fungo do pão *Neurospora crassa*.

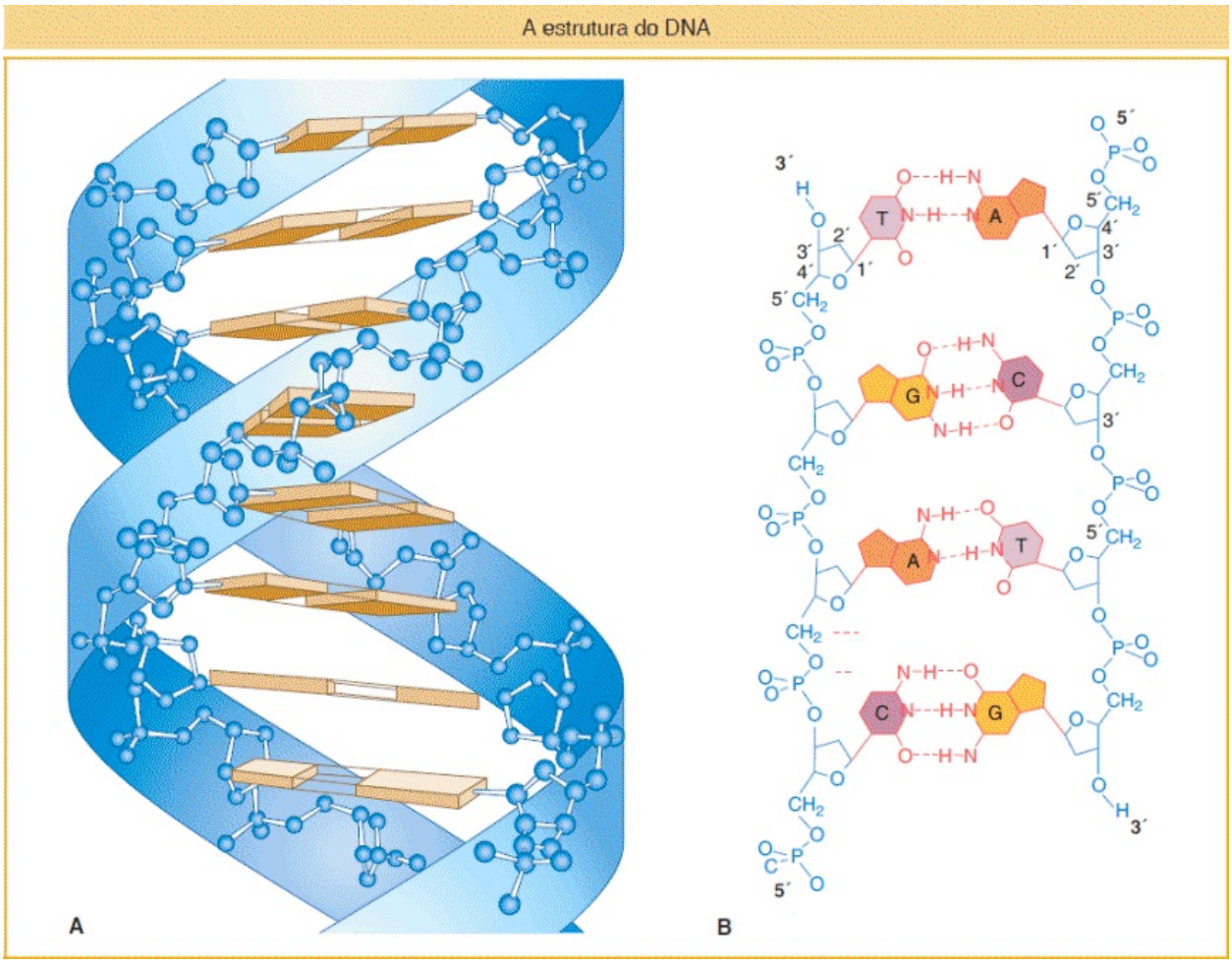


FIGURA 1.8 A. A estrutura da dupla-hélice do DNA, demonstrando os arcabouços de açúcar-fosfato em *azul* e as bases pareadas em *marrom*. **B.** Uma representação plana do DNA, demonstrando como A sempre pareia com T, e G com C. Cada *fileira de traços* entre as bases representa uma ponte de hidrogênio.

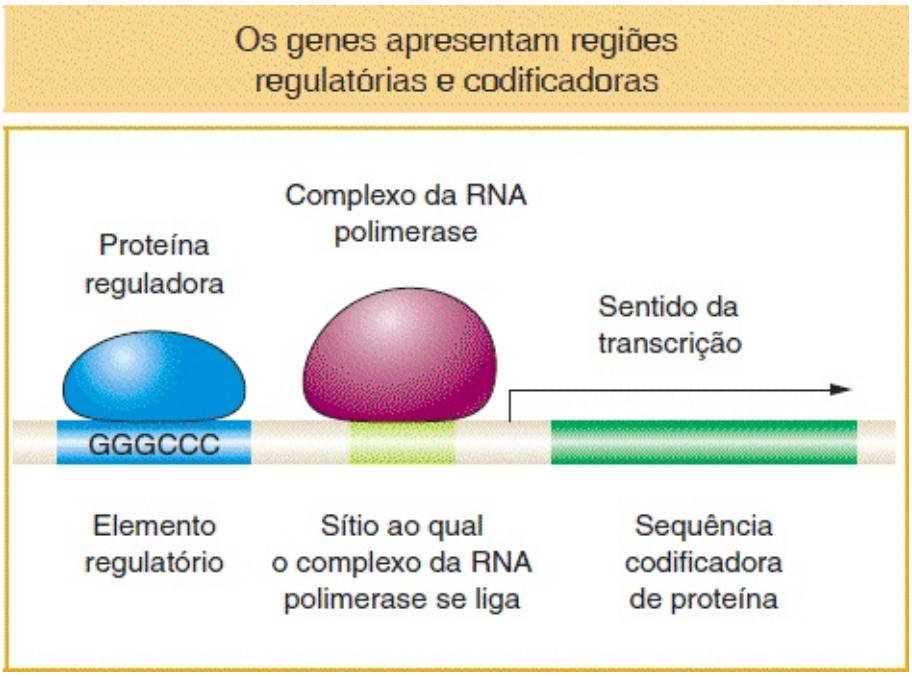


FIGURA 1.9 Estrutura de um gene codificador de proteína, demonstrando um elemento regulatório no DNA (GGGCCC) ao qual uma proteína reguladora se liga, a região promotora na qual o complexo da RNA polimerase se liga para iniciar a transcrição e uma região codificadora de proteína.

> CONCEITO-CHAVE A redescoberta das leis de Mendel lançou uma nova era, na qual os geneticistas resolveram muitas questões fundamentais a respeito da natureza do gene e do fluxo da informação genética dentro das células. Durante essa era, os geneticistas aprenderam que os genes estão localizados em cromossomos e são feitos de DNA. Os genes codificam proteínas, que conduzem o trabalho enzimático básico dentro das células.

Dogma central da biologia molecular

Em 1958, Francis Crick introduziu a expressão “dogma central” para representar o fluxo da informação genética dentro das células, a partir do DNA e daí para o RNA e para a proteína e desenhou um diagrama simples para resumir essas relações (Figura 1.10 A). Curiosamente, Crick escolheu a palavra *dogma* pensando que ela significava “hipótese”, o que era a sua intenção, sem o conhecimento de que o seu real significado é “uma crença que deve ser aceita sem

dúvidas". Apesar desse pensamento inadequado, a expressão apresentava um poder inegável e sobreviveu.

A Figura 1.10 B mostra uma grande parte do que foi aprendido a respeito da bioquímica da herança de 1905 até 1967. Revisemos a riqueza do conhecimento que essa figura simples captura. À esquerda, você visualiza o DNA e uma seta circular, que representa a **replicação do DNA**, o processo por meio do qual uma cópia do DNA é produzida. Esse processo possibilita que cada uma das duas células-filhas que resultam da divisão celular apresente uma cópia completa de todo o DNA da célula genitora. No Capítulo 7, você explorará os detalhes da estrutura do DNA e sua replicação.

Outra seta conecta o DNA ao RNA, simbolizando como a sequência dos pares de bases em um gene (DNA) é copiada para uma molécula de RNA. O processo de síntese do RNA a partir de um molde de DNA é denominado **transcrição**. Uma classe de moléculas de RNA sintetizada por meio da transcrição é o **RNA mensageiro**, ou **mRNA**, abreviadamente. O mRNA é o molde para a síntese proteica. No Capítulo 8, você descobrirá como a transcrição é realizada.

A seta final na Figura 1.10 B conecta o mRNA e a proteína. Tal seta simboliza a síntese proteica ou a **tradução** da informação contida na sequência de bases específica no mRNA em uma sequência de aminoácidos que compõem uma proteína. As proteínas são as oficinas de produção das células e compreendem as enzimas, os componentes estruturais das células e as moléculas para a sinalização celular. O processo de tradução ocorre nos ribossomos no citoplasma de cada célula. No Capítulo 9, você aprenderá como o código genético é escrito em trincas de letras, denominadas **códons**. Um códon é um conjunto de três nucleotídios consecutivos no mRNA, conjunto esse que especifica um aminoácido em uma proteína. CGC especifica o aminoácido arginina; AGC, a serina; e assim por diante.

Desde que Crick propôs o dogma central, vias adicionais do fluxo da informação genética foram descobertas. Atualmente, sabemos que existem classes de RNA que não codificam proteínas, além de casos nos quais o mRNA é editado após a transcrição e outros em que a informação no RNA é copiada de volta para o DNA (ver Capítulos 8, 9 e 15).

1.2 Após decifrar o código

Com as leis básicas da herança amplamente descobertas até o fim da década de 1960, floresceu uma nova era de aplicação da análise genética para um amplo espectro de questões biológicas. Para esse fim, muitos esforços foram e continuam a ser investidos no desenvolvimento dos recursos e das ferramentas para abordar essas questões. Os geneticistas enfocaram suas pesquisas em uma pequena quantidade de espécies, conhecidas como "organismos-modelo", que são bem-adequadas para a análise genética. Eles também desenvolveram uma variedade impressionante de ferramentas para a manipulação e a análise do DNA.

Organismos-modelo

Geneticistas fazem uso em especial de um pequeno conjunto de organismos-modelo para a análise genética. Um **organismo-modelo** é uma espécie utilizada em biologia experimental com a presunção de que o que é aprendido a partir da análise daquela espécie permanecerá verdadeiro em relação às demais, em especial outras que sejam relacionadas de modo próximo. A filosofia subjacente à utilização dos organismos-modelo em biologia foi ironicamente expressada por Jacques Monod: "Qualquer coisa que se observe ser verdadeira para *E. coli* também tem de ser verdadeira para os elefantes."[1]

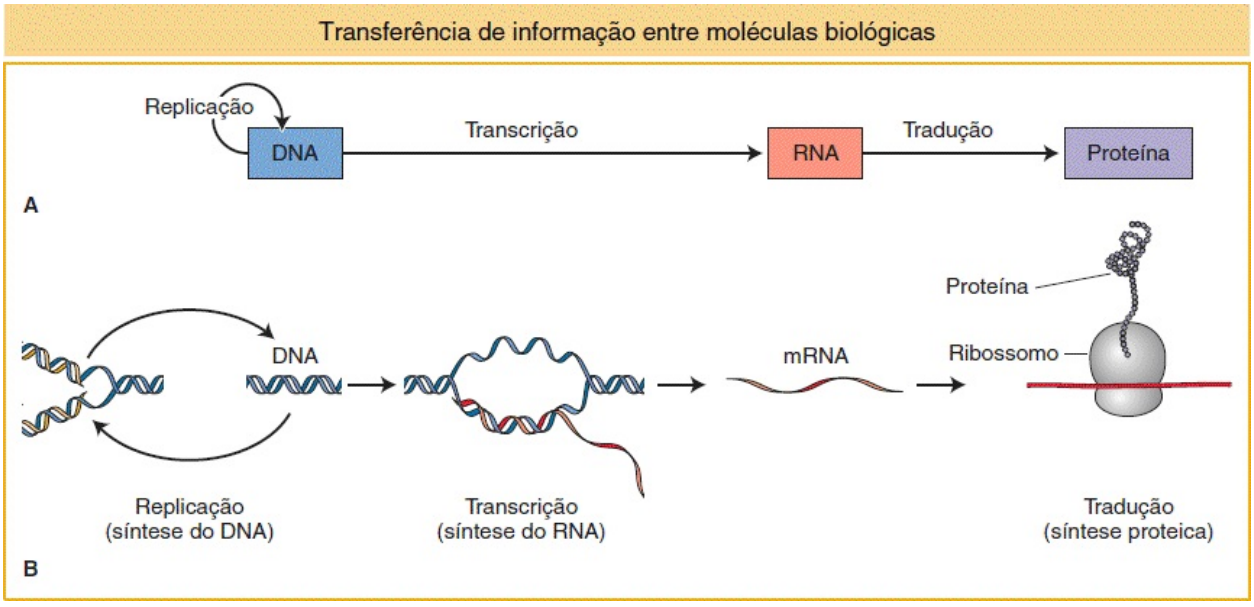


FIGURA 1.10 A. Versão do esboço feito por Francis Crick do dogma central, demonstrando o fluxo de informações entre as moléculas biológicas. A *seta circular* representa a replicação do DNA, a *seta reta* central representa a transcrição do DNA em RNA e a *seta à direita*, a tradução do RNA em proteína. **B.** Esboço mais detalhado, demonstrando como os dois filamentos da dupla-hélice de DNA são replicados de modo independente, como são desassociados para a transcrição e como o RNA mensageiro (mRNA) é traduzido em uma proteína no ribossomo.

Na medida em que a genética amadureceu e enfocou nos organismos-modelo, as ervilhas de Mendel caíram no esquecimento, mas as moscas-das-frutas de Morgan aumentaram em proeminência até se tornarem um dos organismos-modelo mais importantes para a pesquisa genética. Novas espécies foram adicionadas à lista. Uma pequena planta, que cresce como uma erva daninha, denominada *Arabidopsis thaliana* tornou-se a espécie de planta-modelo, e um nematódeo mínimo denominado *Caenorhabditis elegans*, que vive em pilhas de compostagem, tornou-se uma estrela da análise genética na biologia do desenvolvimento (Figura 1.11).

Quais características tornam uma espécie adequada como um organismo-modelo? (1) Pequenos organismos que sejam de manutenção fácil e não dispendiosa são muito convenientes para pesquisas. Assim, moscas-das-frutas são boas, mas baleias-azuis nem tanto. (2) Um curto tempo de geração é imperativo, tendo em vista que os geneticistas, tais como Mendel, precisam cruzar diferentes linhagens e, em seguida, estudar seus híbridos de primeira e segunda gerações. Quanto mais curto o tempo de geração, mais cedo os experimentos podem ser concluídos. (3) Um pequeno genoma é útil. Conforme você aprenderá no Capítulo 15, algumas espécies apresentam genomas grandes, outras, genomas pequenos em termos do número total de pares de bases do DNA. Boa parte do tamanho extra das espécies de genoma grande é composta por elementos repetitivos de DNA entre os genes. Se um geneticista estiver procurando genes, eles podem ser mais facilmente encontrados em organismos com genomas menores e menos elementos repetitivos. (4) Organismos que sejam de fácil cruzamento ou acasalamento e que produzam grandes quantidades de descendentes são os melhores.

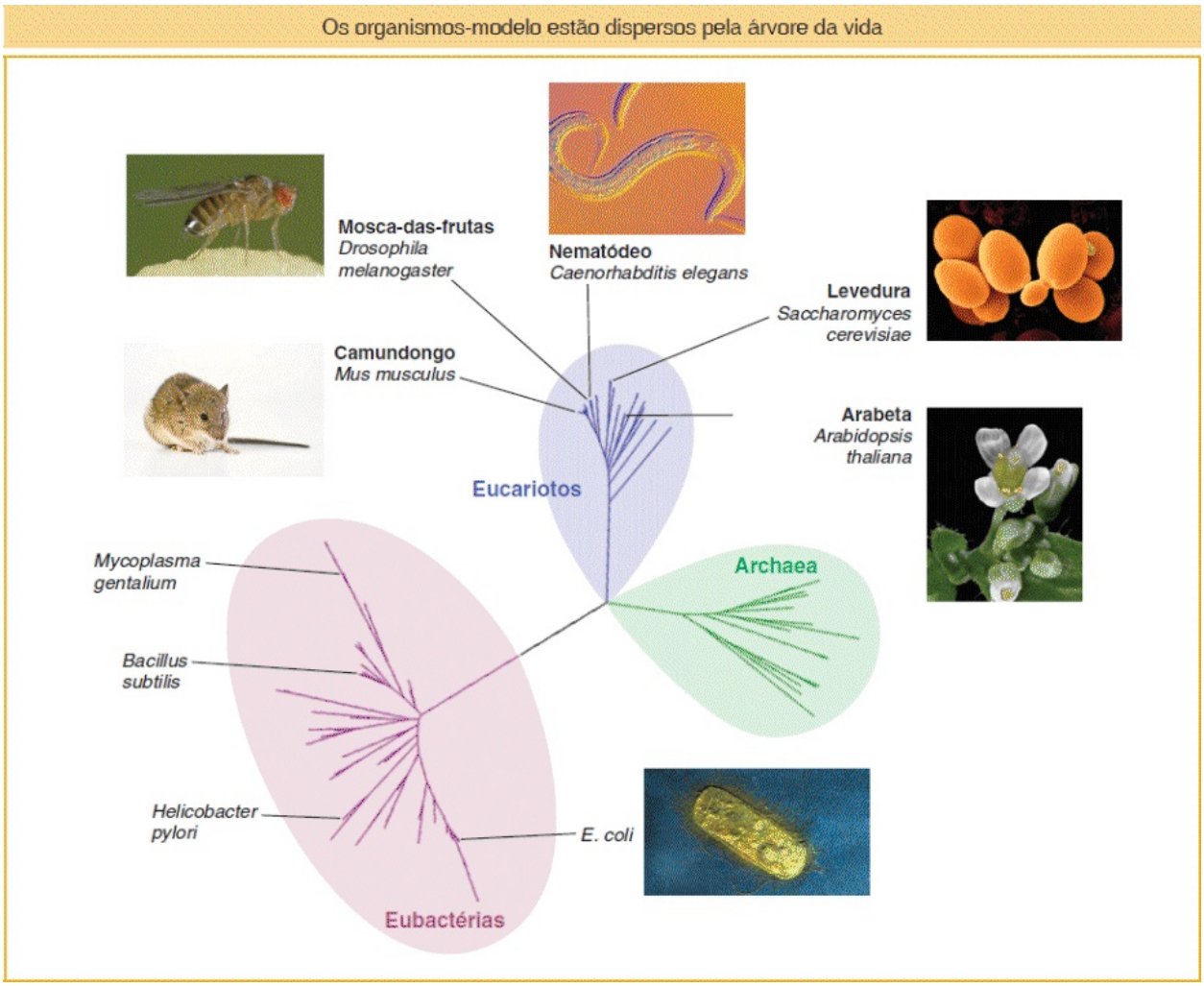


FIGURA 1.11 A árvore demonstra as relações evolutivas entre os principais grupos de organismos: bactérias, *Archaea* e eucariotos (plantas, fungos e animais). (*Em sentido horário, a partir da parte superior, ao centro*: Sinclair Stammers/Science Source; SciMAT/Science Source; Darwin Dale/Science Source; Biophoto Associates/Science Photo Library; Imagebroker.net/SuperStock; ©blickwinkel/Alamy.)

Na medida em que ler este compêndio, você encontrará determinados organismos muitas e muitas vezes. Organismos tais como *Escherichia coli* (uma bactéria), *Saccharomyces cerevisiae* (fermento biológico), *Caenorhabditis elegans* (nematódeo), *Drosophila melanogaster* (mosca-das-frutas) e *Mus musculus* (camundongo) têm sido utilizados repetidamente em experimentos e revelaram muito do que sabemos a respeito de como a herança atua. Os organismos-modelo podem ser encontrados em diversos ramos da árvore da vida (ver Figura 1.11), representando bactérias, fungos, algas, plantas e animais

invertebrados e vertebrados. Essa diversidade possibilita que cada geneticista utilize um modelo mais bem-adequado para uma questão em particular. Cada organismo-modelo apresenta uma comunidade de cientistas que trabalham com ele, que compartilham informações e recursos, facilitando, assim, as pesquisas.

Os experimentos de Mendel foram possíveis porque ele tinha diferentes variedades de ervilha, cada qual carregando uma variante genética diferente em relação a traços, tais como flores roxas *versus* brancas, sementes verdes *versus* amarelas ou caules altos *versus* anões. Para cada uma das espécies-modelo, os geneticistas reuniram muitas variedades (também denominadas cepas ou estoques) com características genéticas especiais que as tornam úteis em pesquisas. Existem cepas de moscas-das-frutas que apresentam variantes de traços, tais como olhos vermelhos *versus* brancos. Existem cepas de camundongos que são propensas ao desenvolvimento de tipos específicos de câncer ou outras condições clínicas, tais como diabetes. Em relação ao fermento biológico, existe uma coleção de quase 5.000 *estoques com deleção*, cada um deles apresentando apenas um gene deletado do genoma. Esses estoques possibilitam que os geneticistas estudem a função de cada gene ao examinar como a levedura é afetada quando o gene é removido. Tendo em vista que *S. cerevisiae* apresenta aproximadamente 6.000 genes no total, essa coleção de 5.000 estoques com deleção abrange a maior parte dos genes no genoma.

As diferentes linhagens de cada organismo-modelo estão disponíveis para os pesquisadores por meio de centros de estoque que mantêm e distribuem as linhagens. As listas dos estoques disponíveis encontram-se na Internet (ver Apêndice B). Para visualizar um exemplo em relação aos estoques de camundongos, dirija-se ao *link* http://jaxmice.jax.org/. Em seguida, clique em "Find JAX mice" na parte superior da página, insira a palavra "black" no campo de pesquisa e clique em "Search". Agora, clique no *link* "C57BL/6J". Você verá uma imagem e informações sobre uma linhagem de camundongos G57-Black comumente utilizada. Outros termos de pesquisa, tais como "albino" ou "obese", ligarão você a linhagens com outras características.

> CONCEITO-CHAVE A maior parte dos estudos genéticos é realizada em um dentre um número limitado de organismos-modelo com características que os tornam especialmente adequados para a análise genética.

Ferramentas para a análise genética

Geneticistas e bioquímicos também criaram uma variedade incrível de ferramentas para a caracterização e a manipulação do DNA, do RNA e de proteínas. Muitas dessas ferramentas estão descritas no Capítulo 10 ou em outros capítulos relevantes para uma ferramenta específica. Existem alguns poucos temas a ser mencionados aqui.

Primeiramente, os geneticistas se aproveitaram do maquinário da própria célula para copiar, colar, cortar e transcrever o DNA, possibilitando que os pesquisadores realizem tais reações dentro de tubos de ensaio. As enzimas que realizam cada uma dessas funções nas células vivas foram purificadas e estão disponíveis para os pesquisadores: as **DNA polimerases** conseguem fazer uma cópia de um único filamento de DNA ao sintetizar um filamento correspondente com a sequência complementar de A, C, G e T. As **nucleases** podem cortar as moléculas de DNA em locais específicos ou degradar toda uma molécula de DNA em nucleotídios únicos. As **ligases** podem unir duas moléculas de DNA a partir de suas extremidades. Com a utilização da DNA polimerase ou de outras enzimas, o DNA também pode ser "rotulado" ou "marcado" com um corante fluorescente ou um elemento radioativo, de modo que o DNA possa ser detectado com a utilização de um detector de fluorescência ou radiação.

Em segundo lugar, os geneticistas desenvolveram métodos para *clonar o DNA* e os genes que ele codifica. Aqui, a clonagem se refere à obtenção de muitas cópias (*clones*) de uma molécula de DNA. O modo comum de fazer isso envolve o isolamento de uma molécula de DNA relativamente pequena (até o comprimento de alguns milhares de pares de bases) a partir de um organismo de interesse. A molécula de DNA pode ser todo um gene ou uma parte dele. A molécula é inserida em um organismo hospedeiro (com frequência *E. coli*), no qual ela é replicada muitas vezes pela DNA polimerase do hospedeiro. Possuir muitas

cópias de um gene é importante para uma vasta variedade de experimentos utilizados para a sua caracterização e manipulação.

Em terceiro lugar, os geneticistas desenvolveram métodos para inserir moléculas de DNA estranhas nos genomas de muitas espécies, incluindo aquelas de todos os organismos-modelo. Esse processo é denominado **transformação** e é possível, por exemplo, transformar os genes de uma espécie no genoma de outra. A espécie receptora se torna, então, um **organismo geneticamente modificado (OGM)**. A Figura 1.12 demonstra uma planta do tabaco, na qual foi inserido um gene do vaga-lume, possibilitando que a planta do tabaco emita luz ou brilhe no escuro.

Em quarto lugar, os geneticistas desenvolveram um grande conjunto de métodos com base na hibridização de moléculas de DNA entre si (ou entre moléculas de RNA). Os dois filamentos complementares de DNA na dupla-hélice são unidos por ligações de hidrogênio, sejam G ≡ C ou A = T. Essas ligações podem ser rompidas por meio de calor (desnaturadas) em uma solução aquosa para fornecer duas moléculas de DNA de filamento único (Figura 1.13). Quando a solução é resfriada sob condições controladas, moléculas de DNA com filamentos complementares hibridizarão preferencialmente entre si. Os métodos de hibridização do DNA possibilitaram muitas descobertas. Por exemplo, o DNA clonado de um gene pode ser marcado com um corante fluorescente e em seguida hibridizado com cromossomos fixados em uma lâmina de microscópio, revelando o cromossomo no qual o gene está localizado.

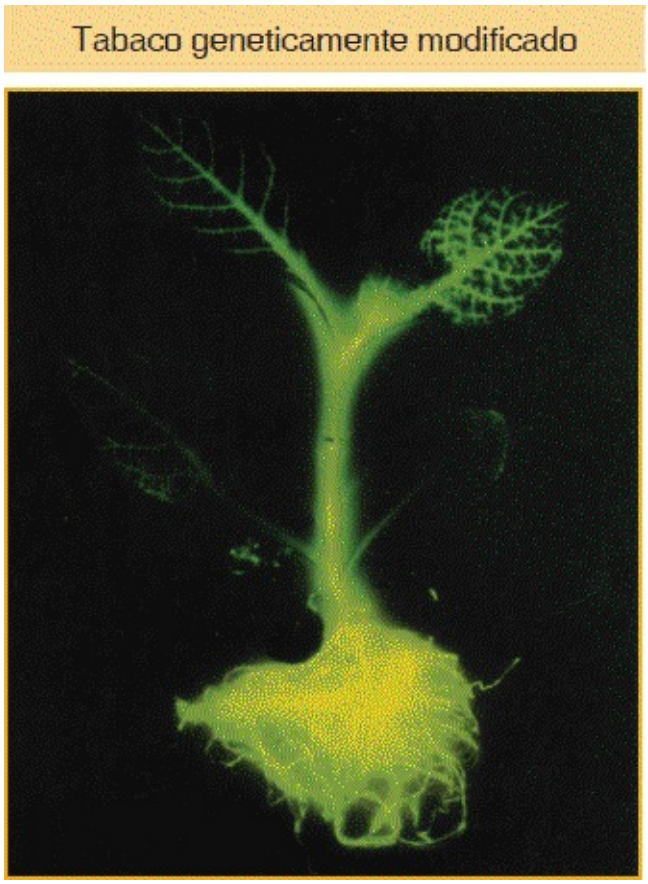


FIGURA 1.12 Esta planta do tabaco geneticamente modificada apresenta um gene do vaga-lume inserido em seu genoma, proporcionando a ela a capacidade de emitir luz. (D. W. Ow *et al.*, "Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants", *Science* 234, 4778, 1986, 856-859.)

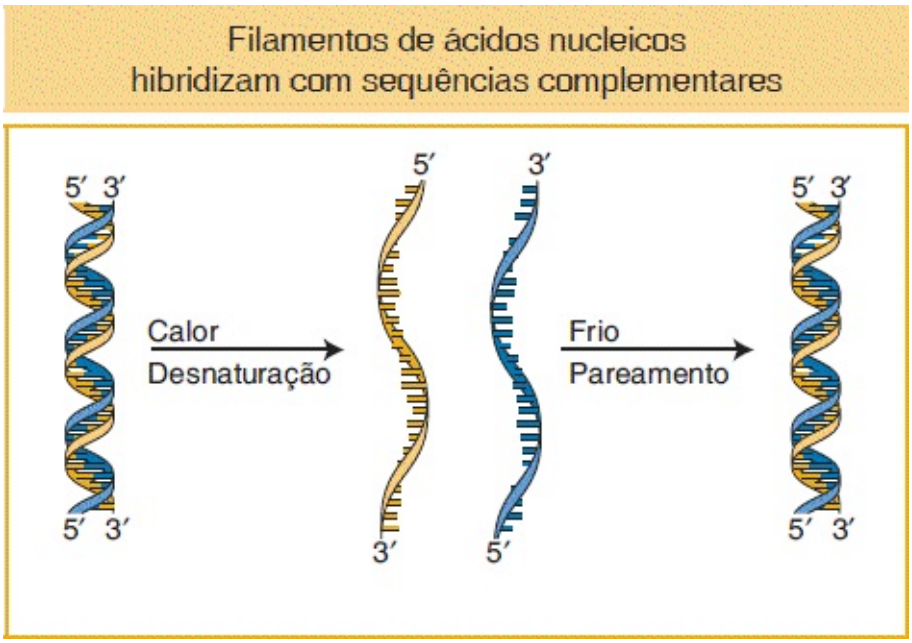


FIGURA 1.13 Os dois filamentos da dupla-hélice do DNA podem ser dissociados por meio de calor em soluções aquosas. Após o resfriamento sob condições controladas, os filamentos se reassociam, ou *hibridizam*, com o seu complemento.

Em quinto lugar, geneticistas e bioquímicos desenvolveram diversos métodos para determinar a sequência exata de todas as A, C, G e T nos genomas, nos cromossomos ou nos genes de um organismo. O processo utilizado para decifrar a sequência exata de A, C, G e T em uma molécula de DNA é denominado **sequenciamento do DNA** e possibilitou que os geneticistas lessem a linguagem da vida.

Finalmente, ao longo dos últimos 20 anos, pesquisadores criaram ferramentas moleculares e matemáticas para a análise de todo o genoma de um organismo em um único experimento. Esses esforços deram origem ao campo da **genômica**: o estudo da estrutura e da função de genomas inteiros (ver Capítulo 14). As ferramentas da genômica possibilitaram que os geneticistas reunissem quantidades extraordinárias de informações sobre os organismos-modelo, incluindo a sequência de DNA completa de seus genomas, listas de todos os seus genes, catálogos de variantes nesses genes, dados sobre os tipos de células e tecidos nos quais cada gene é expresso e muito mais. Para se ter uma ideia do que se encontra disponível, tente navegar no Fly Base (http://flybase.org/), o *site* dos

dados genômicos da mosca-das-frutas (ver Apêndice B).

> CONCEITO-CHAVE O progresso na genética foi produzido e catalisado pelo desenvolvimento de ferramentas moleculares e matemáticas para a análise de genes únicos e genomas inteiros.

1.3 Genética atual

Em uma entrevista em 2008, o geneticista Leonid Kruglyak, da Princeton University, comentou:

> "Você tem esse fenômeno claro e tangível, no qual os filhos se assemelham aos seus genitores. Apesar do que se diga aos estudantes de ciências na escola elementar, nós simplesmente não sabemos como ele atua."

Embora a observação de Kruglyak aparente ser depreciativa do progresso obtido na compreensão da herança ao longo dos últimos 100 anos, certamente não era essa sua intenção. Em vez disso, seu comentário destaca que, apesar das descobertas de alteração do paradigma dos séculos 19 e 20, enigmas são abundantes na genética e a necessidade de novos pensamentos e novas tecnologias nunca foi maior. Mendel, Morgan, Fisher, Watson, Crick e muitos outros (ver Tabela 1.1) delimitaram o fundamento das leis da herança, mas os detalhes que existem sobre aquele fundamento permanecem obscuros. O 1,80 m de DNA na célula única de um zigoto humano codifica as informações necessárias para transformar aquela célula em um adulto, mas como exatamente isso ocorre é compreendido apenas nos detalhes mais escassos.

Nesta seção, revisaremos quatro avanços recentes na genética: descobertas de bastante importância e interesse geral, que receberam destaque na imprensa. A leitura dessas descobertas irá revelar o poder da genética de responder questões críticas a respeito da vida e destacar como esse conhecimento pode ser aplicado para a abordagem de problemas na sociedade. Esse compêndio e o progresso do estudo no qual você está envolvido devem transmitir uma mensagem dupla: a

ciência da genética alterou profundamente a nossa compreensão sobre a vida, mas também é um campo jovem no meio de uma fase dinâmica do seu desenvolvimento.

Da genética clássica à genômica médica

Conheça a paciente VI-1 (Figura 1.14 A). Seu nome é Louise Benge e, quando jovem, ela desenvolveu uma doença incapacitante. No início de seus 20 anos, ela começou a sentir dores insuportáveis nas pernas após caminhar por tão pouco quanto um quarteirão. Primeiramente, ela ignorou a dor; em seguida, conversou com seu médico de assistência primária. Por fim, visitou uma longa lista de especialistas. Ela realizou uma bateria de exames e radiografias, que lhe revelaram o problema: suas artérias, desde a aorta até os membros inferiores estavam calcificadas, obstruídas com depósitos de fosfato de cálcio (Figura 1.14 B). Era uma doença para a qual seus médicos não tinham nome nem tratamento. Ela apresentava uma doença, mas não um diagnóstico. Havia apenas uma coisa a fazer: seu médico encaminhou Benge ao Programa de Doenças Não Diagnosticadas (em inglês, *Undiagnosed Diseases Program* [UDP]) nos National Institutes of Health, em Bethesda, Maryland.

O UDP é um grupo de médicos e cientistas que possuem conexões com especialistas pelos National Institutes of Health, em todos os campos imagináveis da medicina. Essa equipe é solicitada a abordar os casos mais desafiadores. Trabalhando com Benge, a equipe do UDP a submeteu a quase todos os exames em seu arsenal e logo descobriu o defeito subjacente da doença. Benge apresentava um nível muito baixo de uma enzima denominada CD73, a qual está envolvida na sinalização celular e é a responsável por enviar um sinal que bloqueia a calcificação. Com isso, os médicos do UDP poderiam fornecer a Benge um diagnóstico. Denominaram a doença “calcificação arterial em virtude de deficiência de CD73”, ou ACDC [do inglês, *arterial calcification due to deficiency of CD73*].

O que intrigava a equipe do UDP a respeito do caso de Benge era que ela não era a única a apresentar tal doença. Benge tinha dois irmãos e duas irmãs e todos eles apresentavam calcificação arterial. Entretanto, o que era digno de nota era

que os pais de Benge não eram afetados. Além disso, Benge e todos os seus irmãos tinham filhos, mas nenhum apresentava calcificação arterial. Esse padrão de herança sugeria que a causa subjacente fosse genética. Especificamente, ele sugeria que Benge e todos os seus irmãos herdaram duas cópias defeituosas do CD73 ou de um gene que influencia a expressão de CD73 — uma de sua mãe e uma de seu pai. Uma pessoa com uma cópia funcional e uma cópia defeituosa pode ser normal, mas, se ambas as cópias de uma pessoa forem defeituosas, ocorre a ausência da função que o gene proporciona. A situação é justamente igual à das ervilhas com flores brancas de Mendel. Tendo em vista que o alelo funcional é dominante sobre o alelo não funcional, a ACDC, assim como as flores brancas, aparece apenas se um indivíduo carrear dois alelos defeituosos.

A equipe do UDP esmiuçou ainda mais o histórico familiar de Benge e descobriu que seus genitores eram primos em terceiro grau (Figura 1.15). Tal revelação se adequava bem à ideia de que a causa era um gene defeituoso. Quando um marido e uma esposa são parentes próximos, tais como primos em terceiro grau, existe um aumento da chance de que ambos tenham herdado a mesma versão de um gene defeituoso de seu ancestral em comum e de que ambos venham a transmitir esse gene defeituoso aos filhos. Crianças com uma cópia de um gene defeituoso com frequência são normais, mas uma criança que herda uma cópia defeituosa de ambos os genitores provavelmente apresentará um distúrbio genético.

Louise Benge tem uma doença não diagnosticada

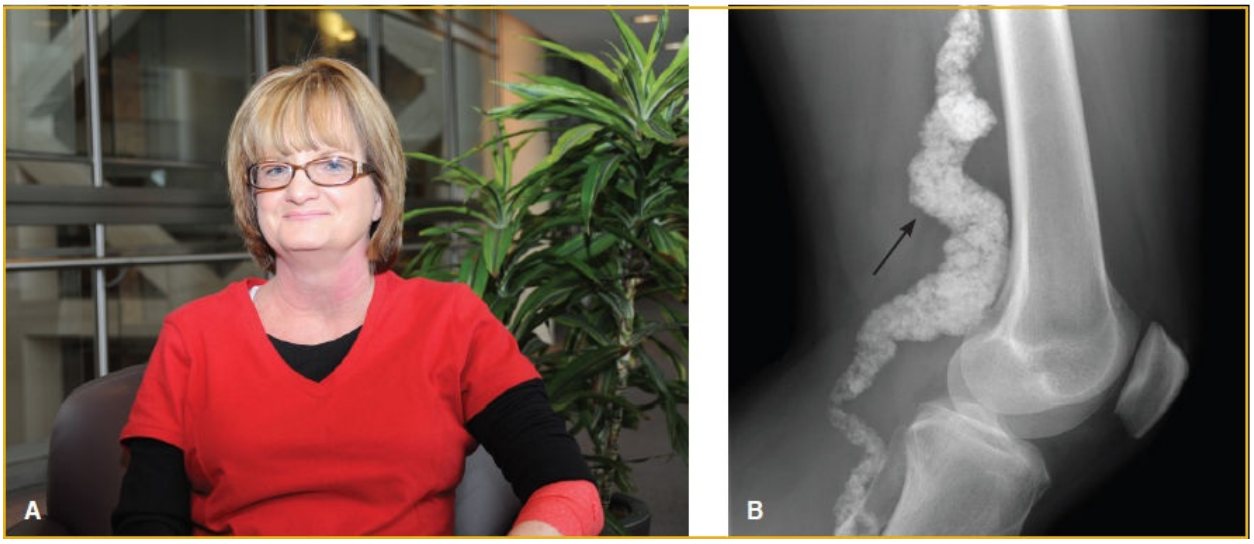


FIGURA 1.14 A. Louise Benge desenvolveu uma doença não diagnosticada quando jovem. **B.** Uma radiografia revelou que a condição da doença de Louise Benge causou a calcificação das artérias em suas pernas. (*A.* Jeannine Mjoseth, NHGRI/www.genome.gov; *B.* National Human Genome Research Institute [NHGRI].)

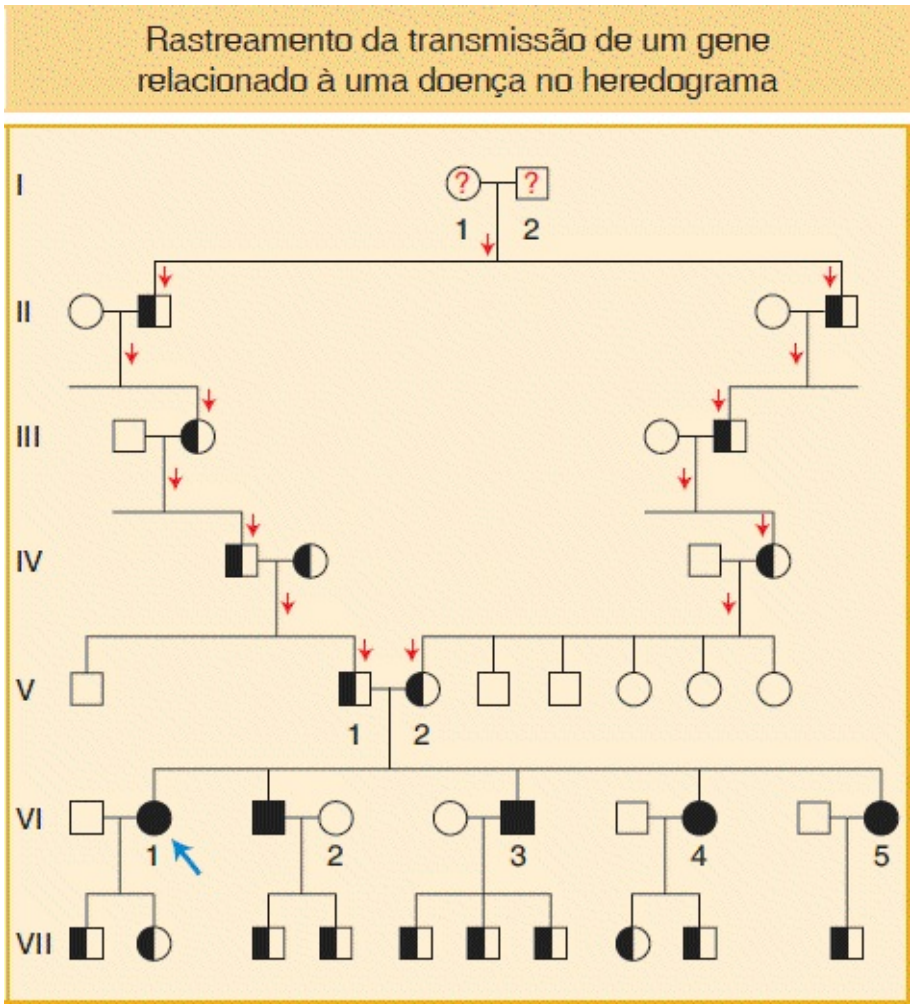


FIGURA 1.15 Árvore genealógica, ou heredograma, demonstrando a herança do gene mutante que causa calcificação arterial em virtude de deficiência de CD73 (ACDC). Os *quadrados* são homens e os *círculos* são mulheres. As *linhas horizontais* que conectam um indivíduo do sexo masculino com um do feminino são os cruzamentos. As *linhas verticais* conectam um casal à sua descendência. Os *numerais romanos* designam as gerações; os *numerais arábicos* designam os indivíduos dentro das gerações. Os *quadrados ou círculos semipreenchidos* indicam um indivíduo que carreia uma cópia do gene mutante. Os *quadrados ou círculos preenchidos* indicam um indivíduo com duas cópias do gene mutante e que apresenta a doença ACDC. O indivíduo I-1 ou I-2 era necessariamente portador do gene mutante, mas qual deles é incerto, conforme indicado por *?*. A *seta azul* indica Louise Benge. As *setas vermelhas* demonstram a transmissão do gene mutante através das gerações. (Dados de C. St. Hilaire *et al.*, *New England Journal of Medicine* 364, 2011, 432-442.)

Na Figura 1.15, podemos observar como isso ocorre. A mãe e o pai de Benge (indivíduos V-1 e V-2 na figura) apresentam os mesmos tataravôs (I-1 e I-2). Se um dos tataravôs era portador de um gene mutante para CD73, ele pode ter sido

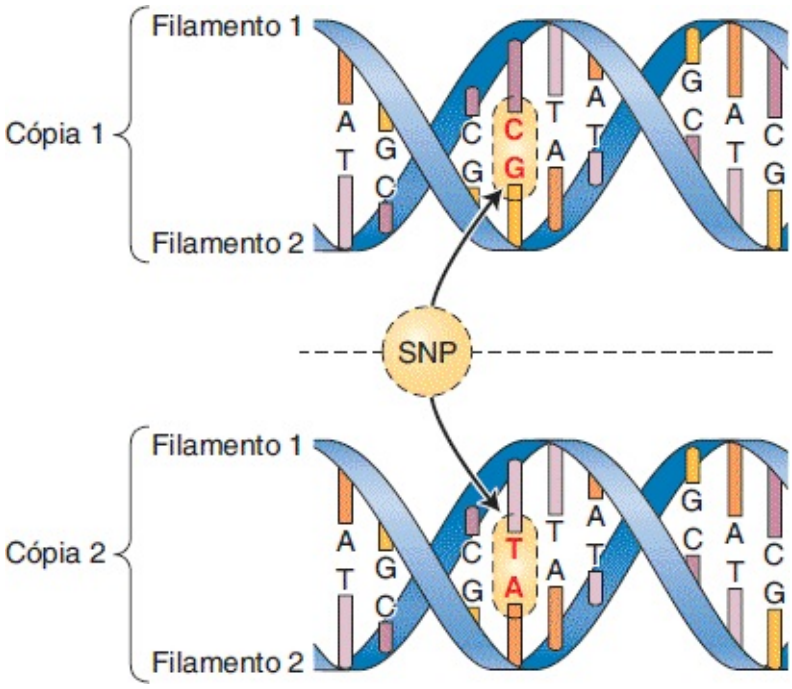
transmitido ao longo das gerações para ambos os genitores de Benge (siga as setas vermelhas). Assim, se Benge recebeu a cópia mutante tanto de sua mãe quanto de seu pai, então ambas as suas cópias serão defeituosas. Cada um de seus irmãos também deve ter herdado duas cópias mutantes de seus genitores, o que explicaria o fato de eles apresentarem ACDC. A chance de algo assim ocorrer é muito pequena. Se ambos os pais de Benge apresentavam uma cópia mutante, a chance de Benge e todos os seus quatro irmãos receberem uma cópia mutante de ambos os genitores é de apenas 1 em 1.024. No Capítulo 2, você aprenderá como calcular essas probabilidades.

Com essa dica do histórico familiar, a equipe do UDP conseguiu descobrir onde procurar o gene mutante no genoma. Eles precisavam procurar um segmento em um dos cromossomos no qual a cópia que Benge herdou de sua mãe fosse idêntica à cópia que ela herdou de seu pai. Além disso, cada um dos irmãos de Benge também deveria apresentar duas cópias desse segmento idênticas às de Benge. Tais regiões são muito raras nas pessoas, exceto se seus genitores forem aparentados, como no caso de Benge, em que seus pais eram primos em terceiro grau. Em geral, um segmento de um cromossomo com algumas centenas de pares de bases de comprimento apresentará diversas diferenças na sequência de A, C, G e T entre a cópia que herdamos de nossa mãe e a que herdamos de nosso pai. Essas diferenças são conhecidas como **polimorfismos de nucleotídio único**, ou, abreviadamente, **SNP** (do inglês, *single nucleotide polymorphisms*; ver Quadro 1.1).

Quadro 1.1 Polimorfismos de nucleotídio único.

Variação genética é qualquer diferença entre duas cópias do mesmo gene ou molécula de DNA. O tipo mais simples de variação genética que se pode observar em um sítio de nucleotídio único é uma diferença na base do nucleotídio presente, seja adenina, citosina, guanina ou timina. Essas variantes são denominadas **polimorfismos de nucleotídio único (SNP)** e são o tipo mais comum de variação na maior parte dos organismos, senão em todos. A figura

demonstra duas cópias de uma molécula de DNA da mesma região de um cromossomo. Observe que as bases são as mesmas nas duas moléculas, exceto onde uma molécula apresenta um par CG e a outra, um par TA. Se lermos o filamento 1 das duas moléculas, a molécula superior apresenta uma "G" e a molécula inferior, uma "A" no sítio do SNP.



A equipe do UDP utilizou uma nova tecnologia genômica, denominada microarranjo de DNA (ver Capítulo 18), que lhe possibilitou o estudo de um milhão de posições de pares de bases em todo o genoma. Em cada uma dessas posições ao longo dos cromossomos, a equipe pôde verificar onde os dois segmentos cromossômicos de Benge eram idênticos e se todos os irmãos de Benge também carreavam duas cópias idênticas daquele segmento. Em relação a Benge, espera-se que uma parte de apenas 1/512 do seu genoma apresente duas cópias idênticas e a chance de que todos os seus quatro irmãos também apresentem as mesmas duas cópias idênticas é ainda menor.

Ao observar os dados de SNP de todo o genoma, a equipe do UDP encontrou exatamente o tipo de segmento cromossômico que estava procurando. Havia um

pequeno segmento em um dos cromossomos de Benge em relação ao qual ela e todos os seus irmãos apresentavam as mesmas duas cópias idênticas. Além disso, eles descobriram que o gene que codifica a enzima CD73 está localizado neste segmento. Esse resultado sugeria que Benge e todos os seus irmãos apresentavam duas cópias idênticas do mesmo gene codificador de CD73 defeituoso. A equipe parecia ter encontrado a agulha que estavam procurando no palheiro; entretanto, havia um último experimento a ser realizado.

A equipe precisava identificar o defeito específico no gene CD73 defeituoso que Benge e seus irmãos haviam herdado. Após a determinação da sequência do DNA em relação ao gene CD73 de Benge e seus irmãos, a equipe encontrou o defeito no gene — "a prova concreta". O gene defeituoso codificava apenas uma proteína curta, ou truncada — ele não codificava a sequência de aminoácidos completa. Um dos códons do DNA com as letras TCG, que codifica o aminoácido serina, havia mutado para TAG, que sinaliza o truncamento da proteína. A proteína sintetizada a partir da versão de Benge do gene CD73 estava truncada, de modo que ela não conseguia sinalizar para manter a via da calcificação desligada nas células das artérias.

A jornada de Louise Benge, desde a primeira dor nas pernas até o conhecimento de que apresentava uma nova doença, denominada ACDC, foi longa. O diagnóstico da doença foi um triunfo possibilitado pela integração da genética de transmissão clássica e da genômica. Conhecer o defeito subjacente à ACDC possibilitou que os médicos tentassem um medicamento que jamais teriam considerado sem que soubessem que a causa era uma enzima CD73 defeituosa, medicamento esse denominado etidronato, o qual pode substituir a CD73 na sinalização celular para manter a via da calcificação desligada. Estudos clínicos com etidronato atualmente estão em andamento para pacientes com ACDC e sua conclusão está programada para 2017.

> CONCEITO-CHAVE A genética de transmissão clássica fornece o fundamento para a genética médica moderna. A integração da genética clássica e das tecnologias genômicas possibilita que as causas de doenças

> herdadas sejam prontamente identificadas.

Investigação da mutação e do risco de doença

Logo após a redescoberta do trabalho de Mendel, o médico alemão Wilhelm Weinberg relatou que parece haver uma incidência mais alta de nanismo com membros curtos (acondroplasia) entre crianças nascidas por último em famílias alemãs do que entre aquelas nascidas primeiro. Algumas décadas mais tarde, o geneticista britânico J. B. S. Haldane observou outro padrão incomum de herança. As genealogias de algumas famílias britânicas sugeriam que novas mutações em relação ao distúrbio da coagulação sanguínea, hemofilia, tendiam a surgir em homens com mais frequência do que em mulheres. Consideradas em conjunto, essas duas observações sugeriram que o risco de uma criança herdar um distúrbio é maior na medida em que os genitores envelhecem e também que os pais apresentam maior probabilidade do que as mães de contribuir com novas mutações para seus filhos.

Ao longo das décadas seguintes, as observações de Weinberg e Haldane foram amparadas por outros estudos, mas os dados não eram conclusivos. Imputar uma nova mutação em uma criança ao pai ou à mãe era uma tarefa difícil e havia uma escassez de famílias bem-adequadas para o estudo da ligação entre a idade parental e novas mutações associadas a doenças. Esses fatores evitaram conclusões definitivas sobre a relação entre a idade parental e a ocorrência de novas mutações.

Em 2012, avanços na genômica e na tecnologia de sequenciamento do DNA (ver Capítulo 14) possibilitaram novas análises, que comprovaram que as suspeitas de Weinberg e Haldane estavam corretas e que forneceram um quadro muito detalhado da origem de novas mutações dentro das famílias. Aqui se encontra como isso foi possível. Uma equipe de geneticistas da Islândia estudou 78 “trios”: grupos familiares compostos por mãe, pai e um filho (Figura 1.16). Os pesquisadores possuíam dados de até três gerações de algumas das famílias, dados esses que correspondiam ao filho, seus genitores e no mínimo um conjunto de avós. Os estudiosos determinaram a sequência completa do genoma de cada

indivíduo com o DNA isolado a partir de suas células sanguíneas, compilando as sequências do genoma de um total de 219 indivíduos. Tendo em vista que cada indivíduo possui duas cópias de cada cromossomo (*i. e.*, duas cópias do genoma humano), seus dados de fato incluem as sequências de 438 genomas.

Com essas sequências do genoma em mãos, os pesquisadores conseguiram vasculhar os dados à procura de *mutações novas* ou *mutações de novo*: variantes de DNA únicas que existem em uma criança, mas em nenhum de seus genitores. Seu enfoque foi nas **mutações de ponto (mutações pontuais)** ou uma alteração de uma letra no código do DNA para outra, que pode ocorrer durante a replicação do DNA (ver Capítulo 16). Por exemplo, uma alteração de uma adenina (A) para uma guanina (G) (Figura 1.17).

A lógica do processo de descoberta utilizada pelos geneticistas islandeses está resumida na Figura 1.17, que demonstra um segmento de DNA para cada membro de um trio. Cada indivíduo apresenta duas cópias do segmento. Observe que a cópia M1 na mãe possui um SNP (*letra verde*) que a distingue da cópia M2. De modo semelhante, existem dois SNP (*letras roxas*) que distinguem as duas cópias do mesmo segmento no pai. Comparando o filho aos genitores, observamos que o filho herdou a cópia M1 de sua mãe e a cópia F2 de seu pai. Observe com mais cuidado as duas cópias do segmento do filho e você verá mais uma coisa. Existe uma variante única (*letra vermelha*) que ocorre no filho, mas em nenhum de seus genitores. Essa é uma mutação pontual *de novo*. Nesse caso, trata-se de uma mutação de uma guanina (G) para uma timina (T). Podemos observar que a mutação surgiu no pai, tendo em vista que se encontra na cópia F2 do segmento.

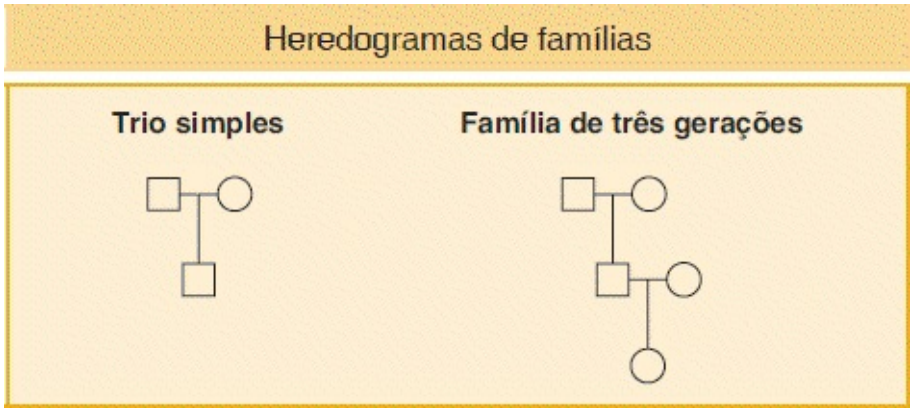


FIGURA 1.16 Os *quadrados* são homens e os *círculos* mulheres. As *linhas horizontais* indicam um cruzamento. As *linhas verticais* conectam um par cruzado à sua descendência.

Onde e exatamente quando a nova mutação ilustrada na Figura 1.17 surgiu? A maior parte de nossos corpos é composta por células somáticas, que constituem tudo, desde o nosso cérebro até o nosso sangue. Entretanto, também apresentamos uma linhagem de células especiais, denominada linhagem germinativa, que se divide para produzir ovócitos nas mulheres e espermatozoides nos homens. As novas mutações, que têm origem nas células somáticas, na medida em que elas se dividem durante o crescimento e o desenvolvimento de nossos corpos, não são transmitidas para nossa descendência. Entretanto, uma nova mutação que ocorre na linhagem germinativa pode ser transmitida para a descendência. A mutação ilustrada na Figura 1.17 surgiu na linhagem germinativa do pai.

Com os dados da sequência genômica em relação aos trios, os geneticistas islandeses fizeram algumas descobertas bastante surpreendentes. Primeiramente, entre os 78 filhos no estudo, eles observaram um total de 4.933 novas mutações de ponto. Cada filho carreava aproximadamente 63 mutações únicas, que não existiam em seus genitores. A maior parte delas ocorreu em partes do genoma que apresentam apenas uma pequena chance de impor um risco à saúde, mas 62 das 4.933 mutações causaram alterações potencialmente nocivas aos genes, pois alteraram a sequência de aminoácidos da proteína codificada. Em segundo lugar, entre as mutações que poderiam ser atribuídas a um genitor de origem, havia em média 55 do pai para cada 14 da mãe. Os filhos estavam herdando quase quatro vezes mais novas mutações de seus pais do que de suas mães. A equipe islandesa havia confirmado a previsão de Haldane realizada 90 anos antes.

As sequências do genoma também possibilitaram que a equipe testasse a previsão de Weinberg de que a frequência das mutações aumenta com a idade dos genitores. Os pesquisadores tomaram conhecimento das idades da mãe e do pai no momento da concepção para cada trio estudado. Quando investigaram se a frequência das mutações aumentava com a idade da mãe ao controlar a idade do pai, a equipe não encontrou qualquer evidência dessa hipótese. Mães mais velhas não transmitiram mais novas mutações de ponto à sua descendência do que as

mães mais novas. (É sabido que mães mais velhas produzem mais aberrações cromossômicas do que mães mais jovens, tais como uma cópia extra do cromossomo 21, que causa síndrome de Down; ver Capítulo 17.) Em seguida, os estudiosos examinaram a relação entre a mutação e a idade do pai ao controlar a idade da mãe. Nesse ponto, eles encontraram uma relação poderosa. Quanto mais velho o pai, mais alta a frequência de novas mutações de ponto (Figura 1.18). De fato, para cada ano de aumento em sua idade, um pai transmitirá duas novas mutações adicionais aos seus filhos. Um pai de 20 anos de idade transmitirá aproximadamente 25 novas mutações a cada um de seus filhos, mas um pai de 40 anos de idade transmitirá aproximadamente 65 novas mutações. A observação de Weinberg realizada 100 anos antes foi confirmada.

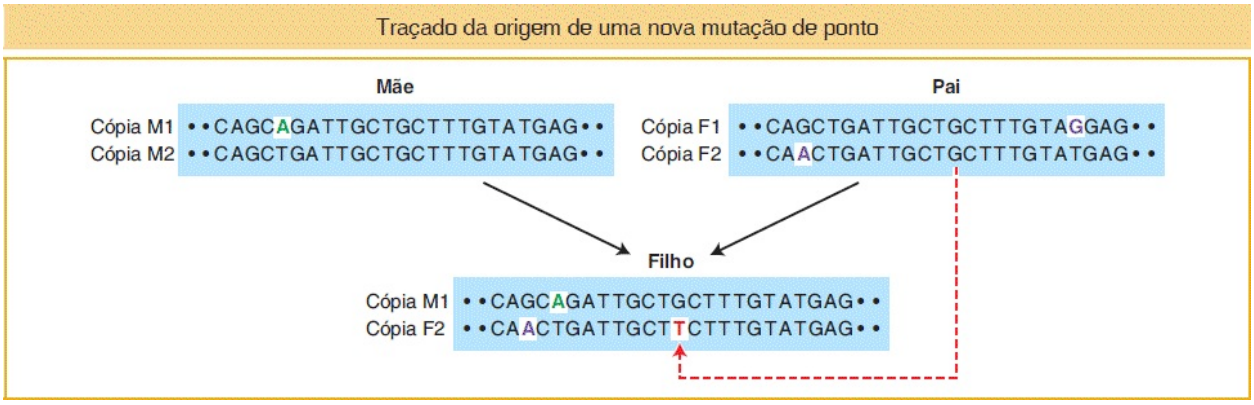


FIGURA 1.17 É demonstrado um segmento curto do DNA de um dos cromossomos. Cada indivíduo apresenta 2 cópias do segmento. Na mãe, elas estão rotuladas M1 e M2; no pai, F1 e F2. O filho herdou a cópia M1 de sua mãe e a F2 de seu pai. A versão de F2 no filho carreia uma nova mutação de ponto (*vermelho*). Os polimorfismos de nucleotídio único (SNP) que distinguem as diferentes cópias estão demonstrados em *verde* (mãe) e em *roxo* (pai).

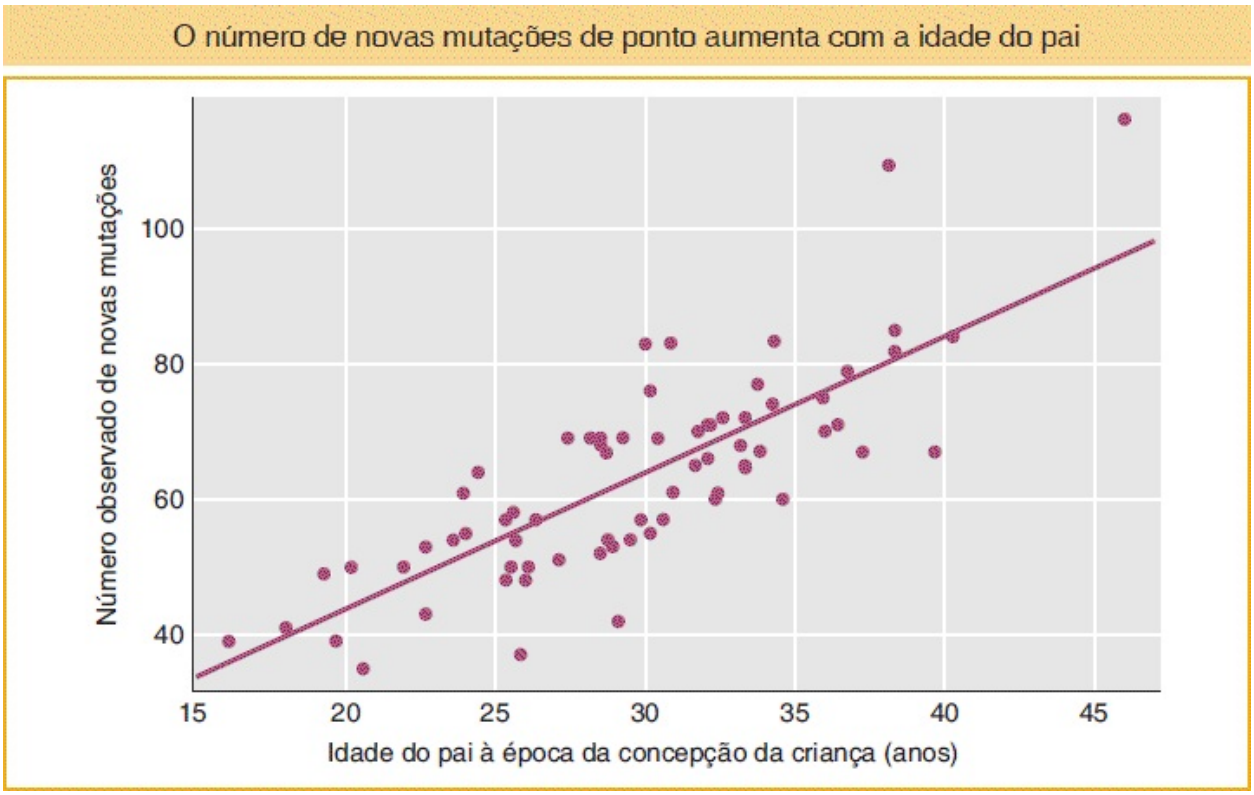


FIGURA 1.18 Gráfico do número de novas mutações de ponto em cada filho (*eixo y*) de acordo com a idade do pai da criança (*eixo x*). Cada ponto representa um dos 78 filhos estudados. A linha diagonal indica a taxa de aumento em novas mutações com a idade do pai. (Dados de A. Kong *et al.*, *Nature* 488, 2012, 471-475.)

Por que a idade do pai é importante enquanto a da mãe aparenta não apresentar efeito sobre a frequência de novas mutações de ponto? A resposta está nos diferentes modos por meio dos quais homens e mulheres formam gametas. Nas mulheres, assim como nas fêmeas de outros mamíferos, o processo de produção dos ovócitos ocorre principalmente antes do nascimento da mulher. Portanto, quando uma mulher nasce, possui em seus ovários um conjunto de células precursoras dos ovócitos, que amadurecerão até ovócitos sem ciclos adicionais de replicação do DNA. Na mulher, a partir do ponto em que ela foi concebida até a formação dos ovócitos em seus ovários, existem aproximadamente 24 ciclos de divisão celular, 23 dos quais apresentam um ciclo de replicação cromossômica (DNA) e uma oportunidade para que ocorra um erro de cópia ou uma mutação. Todos os 23 desses ciclos de replicação cromossômica ocorrem antes do nascimento da mulher, de modo que não existem ciclos adicionais após o seu

nascimento e nenhuma chance de mutações adicionais na medida em que ela envelhece. Portanto, mães mais velhas não contribuem com mais novas mutações de ponto para seus filhos do que mães mais jovens.

A produção de espermatozoides é completamente diferente. As divisões celulares que produzem os espermatozoides continuam durante toda a vida de um homem e existem muito mais ciclos de divisão celular na formação dos espermatozoides do que na formação dos ovócitos. Os espermatozoides produzidos por homens de 20 anos de idade terão sido submetidos a aproximadamente 150 ciclos de replicação do DNA desde o momento da concepção do homem, quase sete vezes mais do que em relação aos ovócitos produzidos por mulheres de 20 anos de idade. Na ocasião em que um homem chega aos 40 anos de idade, seus espermatozoides apresentarão um histórico que envolve mais de 25 vezes mais ciclos de replicação do DNA do que em relação aos ovócitos em uma mulher da mesma idade. Portanto, existe muito mais risco de ocorrência de novas mutações de ponto durante estes ciclos extras de divisão celular e replicação do DNA com o aumento da idade do pai.

Existe uma reviravolta final no projeto extraordinário realizado pelos geneticistas islandeses. Os 78 trios que eles estudaram foram escolhidos em virtude de os filhos, na maior parte dos trios, terem distúrbios hereditários. Foram incluídos 44 filhos com distúrbios do espectro autista e 21 com esquizofrenia. Não havia casos desses distúrbios entre os parentes de nenhum deles, sugerindo que sua condição ocorria em virtude de uma mutação nova. Conforme esperado, os pesquisadores observaram uma correlação entre a idade do pai e o risco de doença — pais mais velhos apresentavam maior probabilidade de ter filhos com autismo e esquizofrenia. Em diversos casos, os dados do DNA da criança e dos genitores também possibilitaram que os pesquisadores identificassem novas mutações específicas em genes que provavelmente causavam o distúrbio. Por exemplo, uma criança com autismo herdou uma nova mutação no gene do *receptor de EPH B2 (EPHB2)*, que atua no sistema nervoso e no qual uma mutação havia sido encontrada anteriormente em uma criança autista.

Estudos como esse podem apresentar importantes implicações para os indivíduos e a sociedade. Alguns homens que pretendem postergar a decisão de

ter filhos poderiam optar por congelar amostras de seus espermatozoides enquanto ainda são jovens. O estudo também nos informa que mudanças na sociedade podem impactar o número de novas mutações que entram no *pool* gênico humano. Se os homens optarem por adiar a paternidade para cursar a faculdade ou o estabelecimento de suas carreiras, haverá um aumento correlato no número de novas mutações entre seus filhos. É de conhecimento comum que a infertilidade aumenta com a idade nas mulheres. Como dito com frequência, o "relógio biológico" de uma mulher começa a funcionar a partir da puberdade. O trabalho dos geneticistas islandeses nos informa que o tempo também corre para os homens.

> CONCEITO-CHAVE As sequências do genoma dos genitores e de seus filhos esclarecem os fatores que contribuem para novas mutações de ponto. Os pais contribuem com quatro vezes mais novas mutações para a sua descendência do que as mães. O número de novas mutações transmitidas de um pai para seus filhos aumenta com a idade do pai.

Quando os pés de arroz ficam um pouco molhados

Entre as plantações de cereais, o arroz é único. Enquanto trigo, cevada, milho e as outras plantações de grãos crescem somente em campos secos, o arroz comumente é cultivado em campos alagados denominados arrozais (Figura 1.19). A capacidade do arroz de crescer em campos alagados oferece a ele uma vantagem: consegue sobreviver a pequenos alagamentos (até 25 cm de água parada) nos arrozais, mas a maioria das ervas daninhas não consegue. Assim, os agricultores de arroz conseguem utilizar os alagamentos para controlar as ervas daninhas em seus campos enquanto seu arroz se desenvolve.

A estratégia funciona bem nos locais em que os agricultores possuem sistemas de irrigação para controlar os níveis de água nas plantações e as chuvas fortes não excedem sua capacidade de controlar esses níveis. Se a água nos arrozais se tornar muito profunda (superior a 50 cm) durante um período prolongado, as

plantas do arroz, assim como as ervas daninhas, podem sofrer e até morrer.

Arroz crescendo em um campo alagado ou arrozal

FIGURA 1.19 O arroz é cultivado em campos com água parada denominados arrozais. O arroz está adaptado para tolerar níveis modestos de água parada, mas a água impede o crescimento de ervas daninhas que poderiam competir com o arroz. (©Dinodia/AGE Fotostock.)

A agricultura de arrozais, conforme praticada nas planícies da Índia, do Sudeste Asiático e da África Ocidental, depende da precipitação atmosférica natural, em vez da irrigação, para alagar os campos. Essa circunstância impõe um risco. Quando as chuvas são fortes, a profundidade da água nos arrozais pode exceder os 50 cm e submergir completamente as plantas, fazendo com que as plantas do arroz sofram uma perda no rendimento ou simplesmente morram. Dos 60 milhões de hectares de arrozais em planícies alimentadas pela chuva, um terço sofre com alagamentos prejudiciais regularmente. Estima-se que as chuvas fortes e as monções que alagam os campos causem uma perda superior a US$ 1 bilhão a cada ano. Apenas na Índia, na Indonésia e em Bangladesh, 4 milhões de toneladas de arroz são perdidas em virtude dos alagamentos a cada ano, o suficiente para alimentar 30 milhões de pessoas. Tendo em vista que essa perda, em sua maior parte, é sofrida pelos fazendeiros mais pobres, ela pode levar à desnutrição e até

à inanição.

No início da década de 1990, David Mackill, um geneticista de plantas e um cultivador no International Rice Research Institute, teve uma ideia sobre como melhorar o arroz de modo que ele pudesse tolerar a submersão. Ele identificou uma extraordinária variedade de arroz denominada FR13A, que conseguia sobreviver submersa e mesmo se desenvolver após as plantas terem permanecido totalmente submersas em água profunda por até 2 semanas. Infelizmente, a FR13A apresentava baixo rendimento e a qualidade de seus grãos era mínima. Então, Mackill estabeleceu a transferência de fator(es) genético(s) da FR13A em relação à tolerância à submersão para uma variedade de arroz com mais alto rendimento e mais alta qualidade de grãos. Primeiramente, cruzou FR13A com uma variedade superior de arroz e em seguida cruzou, durante diversas gerações, as plantas híbridas novamente com a variedade superior, até que criou um tipo de arroz melhor, que combinava a tolerância à submersão e o alto rendimento.

Mackill havia conquistado seu objetivo inicial de transferir a tolerância à submersão para uma variedade superior, mas a base genética de a FR13A ser tolerante à submersão permanecia obscura. A tolerância à submersão da FR13A era controlada por muitos genes em cromossomos múltiplos ou poderia ser controlada principalmente por apenas um gene? Para esmiuçar a base genética da tolerância à submersão, Mackill e sua equipe conduziram um tipo de análise genética denominada mapeamento de ***locus* de traço quantitativo (QTL)** (ver Capítulo 19). Um QTL é um *locus* genético que contribui de modo incremental ou quantitativo para a variação em relação a um traço. O gene de Mendel em relação à cor das flores apresentava dois alelos categóricos: um para as flores roxas e o outro para as flores brancas. O QTL apresenta alelos que normalmente geram apenas alterações parciais, tais como a diferença entre um roxo claro e um roxo médio. Com a utilização do mapeamento do QTL, Mackill compreendeu que o segredo da excepcionalidade da FR13A ocorria, em sua maior parte, por causa de um *locus* genético único ou QTL, em um dos cromossomos do arroz. Ele denominou tal *locus SUB1*, em referência a "tolerância à submersão".

Com a localização cromossômica do *SUB1* revelada, era o momento de esmiuçar ainda mais profundamente e identificar a natureza molecular do *SUB1*.

Qual tipo de proteína ele codificava? Como o alelo de *SUB1* observado na FR13A possibilitava que a planta lidasse com a submersão? Qual é a resposta fisiológica que possibilita que a planta sobreviva à submersão?

Para abordar essas questões, as geneticistas moleculares Pamela Ronald, da California University, Davis e Julia Bailey-Serres, da California University, Riverside, uniram-se à equipe. Trabalhando com Mackill, a equipe, expandida, enfocou no segmento do cromossomo que contém o QTL *SUB1* e determinou que ele engloba um membro de uma classe de genes denominados *fatores de resposta ao etileno (ERF)*. Os genes ERF codificam proteínas regulatórias que se ligam a elementos regulatórios em outros genes e, assim, regulam a sua expressão. Portanto, o *SUB1* é um gene que regula a expressão de outros genes. Além disso, eles determinaram que o alelo do *SUB1* na FR13A é ativado em resposta à submersão, enquanto o alelo do *SUB1* encontrado em variedades sensíveis à submersão não é ativado pela submersão.

A próxima pergunta era, como a ativação do *SUB1* possibilita que a FR13A sobreviva à submersão completa? Para responder a essa questão, revisemos como as plantas do arroz comuns respondem à submersão. Quando uma planta está completamente submersa, os níveis de oxigênio em suas células diminuem até um nível baixo e a concentração de etileno, um hormônio das plantas, aumenta nas células. O etileno sinaliza para a planta que escape à submersão por meio do alongamento de folhas e caules para manter a "cabeça" acima da água. Essa *estratégia de escape* funciona bem, desde que a água não seja profunda a ponto de a planta falhar em crescer o suficiente para posicionar caules e folhas acima do nível de alagamento. Se o alagamento for excessivo, a planta não consegue crescer o suficiente para escapar. Na tentativa de crescer e escapar a um alagamento profundo, a planta utiliza todas as suas reservas energéticas (carboidratos), torna-se longa e fraca e finalmente morre.

Como a variedade FR13A consegue sobreviver à submersão enquanto muitos outros tipos de arroz não conseguem? A FR13A apresenta uma estratégia diferente, que pode ser denominada *esperar pacientemente*. Em resposta à submersão completa, em vez de tentar um crescimento rápido para escapar ao alagamento, uma planta FR13A que utiliza a estratégia de esperar sentada se torna

quiescente. Ela interrompe a resposta de crescimento por alongamento, evitando, assim, o esgotamento de sua reserva de carboidratos e consequente enfraquecimento. Com a estratégia da espera, uma planta consegue permanecer em um estado submerso e quiescente por até 2 semanas, permanecendo saudável para retomar o crescimento normal quando o alagamento retroceder.

A estratégia da espera paciente utilizada pela FR13A é controlada pelo *SUB1*, que atua como o interruptor principal ou gene regulatório, para a ativação desse mecanismo. Quando o alagamento aumenta, a concentração do hormônio etileno também aumenta nas células da planta. Tendo em vista que o *SUB1* é um ERF, ele é ligado em resposta à elevação dos níveis de etileno. Em seguida, a proteína que o *SUB1* codifica orquestra a resposta da planta ao ligar (ou desligar) uma bateria de genes envolvidos no crescimento e no metabolismo da planta. Nas plantas FR13A que se tornam submersas, os genes envolvidos no alongamento do caule e das folhas como parte da estratégia de escape são desligados, assim como os genes envolvidos na mobilização das reservas de energia (carboidratos) necessárias para alimentar a estratégia de escape. Com a utilização de ferramentas de genética molecular e genômica, tais como microarranjos de DNA (ver Capítulos 10 e 14), a equipe do arroz foi capaz de decifrar o extensivo catálogo de genes que controlam o alongamento do órgão, o metabolismo do carbono, a florescência e a fotossíntese, que são regulados pelo *SUB1* para a conquista da resposta de esperar sentada.

Com a genética básica do *SUB1* elucidada, era o momento de colocar em funcionamento esse conhecimento. A equipe repetiu o trabalho de cultivo inicial de Mackill para transferir a tolerância ao alagamento para uma variedade superior. Agora, entretanto, tendo em vista que conheciam a localização precisa do *SUB1* em um dos cromossomos, eles puderam transferi-lo para uma variedade superior com precisão cirúrgica. Essa precisão é importante pois possibilitou que a equipe evitasse a transferência de outros genes indesejáveis ao mesmo tempo. Para esse projeto, os pesquisadores trabalharam com uma variedade indiana intolerante à submersão, porém superior, denominada *Swarna*, que é amplamente cultivada e favorecida pelos fazendeiros. A nova linhagem que criaram é denominada *Swarna-sub1* e ela correspondeu às expectativas. Estudos a campo

demonstraram uma diferença surpreendente na sobrevivência e no rendimento das plantas entre a *Swarna* e a *Swarna-sub1* quando existe uma submersão completa. Conforme demonstrado na Figura 1.20, a *Swarna-sub1* proporciona rendimento superior à *Swarna* original sob todos os diferentes níveis de alagamento. Em diversos estudos, o *SUB1* melhorou o rendimento entre uma a três toneladas de grãos por hectare.

Com o amparo e o patrocínio de organizações de pesquisa internacionais, agências governamentais e filantropias, a *Swarna-sub1* e outras variedades superiores que carreiam o alelo *SUB1* da FR13A agora foram distribuídas para os fazendeiros. Em 2008, apenas 700 fazendeiros estavam cultivando o arroz melhorado com *SUB1*, mas em 2012, esse número havia aumentado para 3,8 milhões de fazendeiros. Em 2014, a quantidade de fazendeiros que cultivam arroz com *SUB1* deveria aumentar para 5 milhões, com adição considerável à segurança alimentar entre alguns dos fazendeiros mais pobres do mundo.

Ao longo prazo, o impacto da pesquisa sobre o *SUB1* pode não estar limitado ao arroz. Muitos cultivos estão sujeitos a alagamentos prejudiciais, que reduzem as produções ou destroem todo o cultivo. A pesquisa genética sobre o *SUB1* proporcionou uma profunda compreensão da genética molecular de como as plantas respondem aos alagamentos. Com esse conhecimento, será possível manipular os genomas de outras plantas cultivadas, de modo que elas também possam suportar que seus pés fiquem um pouco mais molhados.

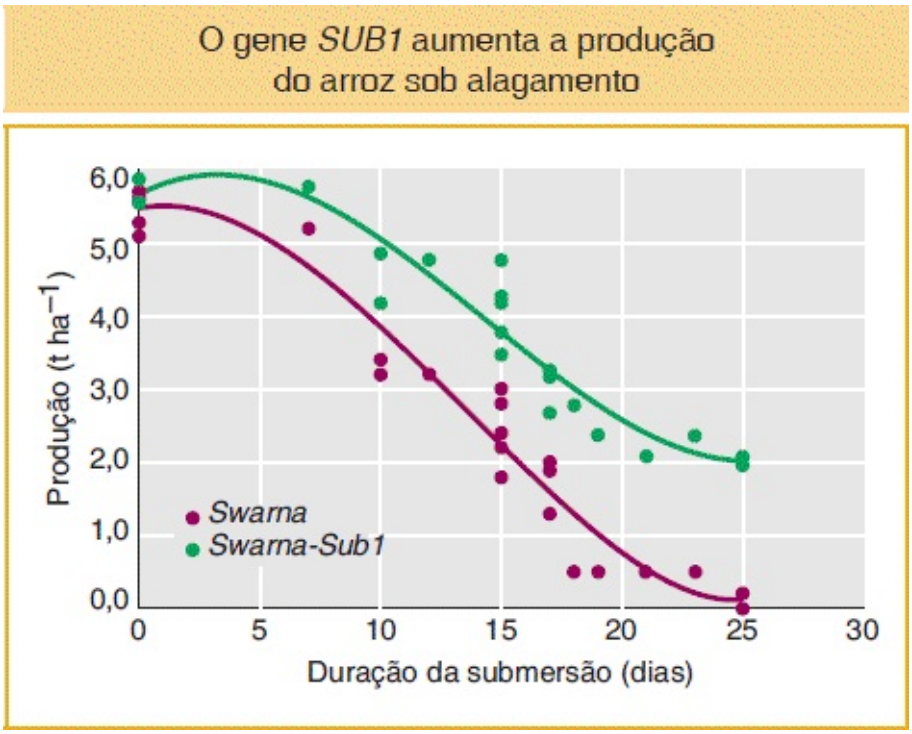


FIGURA 1.20 Comparação do rendimento da variedade *Swarna*, que não é tolerante ao alagamento (*círculos roxos*) e da variedade *Swarna-sub1*, que é tolerante (*círculos verdes*). Rendimento em toneladas por hectare (*eixo y*) *versus* duração do alagamento em dias (*eixo x*). (Dados de Ismail *et al.*, "The contribution of submergence-tolerant (Sub 1) rice varieties to food security in flood-prone rainfed lowland areas in Asia", *Field Crops Research* 152, 2013, 83-93, ©Elsevier.)

CONCEITO-CHAVE A genética e a genômica estão desempenhando um papel de liderança no melhoramento do cultivo de plantas. Os princípios básicos da genética, que você aprenderá durante o seu curso sobre genética, são o fundamento para esses avanços.

Evolução recente em humanos

Um objetivo da genética é compreender as regras que controlam como os genes e as informações que eles codificam são alterados entre as gerações dentro das populações. Os genes nas populações são alterados ao longo do tempo em virtude de diversos motivos diferentes. Por exemplo, conforme vimos, a mutação na linhagem germinativa pode originar uma nova variante gênica ou um novo alelo na

próxima geração, que não estava presente na geração atual. Outro fator é a *seleção natural*, que foi descrita pela primeira vez por Charles Darwin. Em resumo, se os indivíduos com uma determinada variante genética contribuem com mais descendentes para a próxima geração do que indivíduos que não apresentam aquela variante, a frequência daquela variante aumentará ao longo do tempo na população. Os últimos três capítulos do livro enfocam as regras que controlam a transmissão dos genes de uma geração para a próxima dentro das populações.

Ao longo da última década, geneticistas evolutivos descreveram em detalhes extraordinários como as alterações genéticas possibilitaram que as populações humanas se adaptassem às condições de vida em diferentes partes do globo. Esse trabalho revelou que três fatores foram particularmente poderosos na moldagem dos tipos de variantes gênicas que ocorrem em diferentes populações humanas. São eles: (1) patógenos tais como os de malária ou varíola; (2) condições climáticas locais, incluindo radiação solar, temperatura e altitude; e (3) dieta, como as quantidades relativas de carne, cereais ou laticínios ingeridos. No Capítulo 20, você aprenderá como uma variante genética no gene da hemoglobina possibilitou que pessoas na África se adaptassem às devastações da malária. Observemos brevemente exemplos de adaptações genéticas ao clima e à dieta. Daremos início com um caso de adaptação humana à vida em altas altitudes.

Adaptação às altas altitudes. Em seu esforço de colonizar a Cordilheira dos Andes, na América Sul, colonizadores espanhóis estabeleceram cidades no alto das montanhas, próximas aos assentamentos dos povos nativos. Logo perceberam que algo estava errado. Os espanhóis não estavam gerando filhos. Em Potosi, na Bolívia, que está situada 4.000 m acima do nível do mar, 53 anos se passaram após a fundação da cidade antes que nascesse o primeiro filho de espanhóis. Conforme observado pelo padre espanhol Cobo: “Os índios são mais saudáveis e se multiplicam de modo mais prolífero nestas mesmas temperaturas frias, o que é completamente contrário ao que ocorre com os filhos dos espanhóis, cuja maior parte, quando nascida nestas regiões, não sobrevive.”[2] Contrariamente aos nativos andinos, os espanhóis estavam apresentando doença crônica da montanha (DCM), uma condição causada pela sua incapacidade em obter oxigênio suficiente do ar

rarefeito das montanhas.

Desde observações iniciais como essas, geneticistas investiram muitos esforços no estudo da adaptação humana às altas altitudes na América do Sul, no Tibete e na Etiópia. O que possibilita que os nativos dessas regiões prosperem, enquanto nativos de planícies que se mudam para as altas elevações sofrem as graves consequências de saúde da DCM? Vejamos o caso no Tibete, onde os montanheses tibetanos vivem em altitudes de até 4.000 metros acima do nível do mar (Figura 1.21). O alto Platô Tibetano foi colonizado há aproximadamente 3.000 anos e os colonizadores eram relacionados de modo próximo aos chineses modernos Han. Entretanto, em altas altitudes, tibetanos nativos apresentam muito menos probabilidade do que os chineses Han de apresentar DCM e condições como hipertensão pulmonar e a correlata formação de coágulos sanguíneos que a fundamenta.

Para compreender a genética de como os tibetanos se adaptaram à vida em altas elevações, uma equipe de pesquisadores liderada por Cynthia Beall, da Case Western Reserve University, comparou os tibetanos aos chineses Han em mais de 500.000 SNP ao longo do genoma. Tendo em vista que os tibetanos e os chineses são relacionados de modo próximo, espera-se que cada variante de SNP ocorra aproximadamente na mesma frequência em ambos os grupos. Se a variante T de um SNP ocorrer a uma frequência de 10% em chineses Han, ela também deve ser de aproximadamente 10% em tibetanos. Entretanto, se a variante estiver associada à melhora da saúde em altas elevações, a sua frequência teria aumentado entre os tibetanos ao longo das muitas gerações desde que eles colonizaram o Platô Tibetano, tendo em vista que tibetanos com esta variante teriam sido mais saudáveis e teriam tido mais filhos sobreviventes do que aqueles sem a variante. A seleção natural de Charles Darwin estaria atuando.

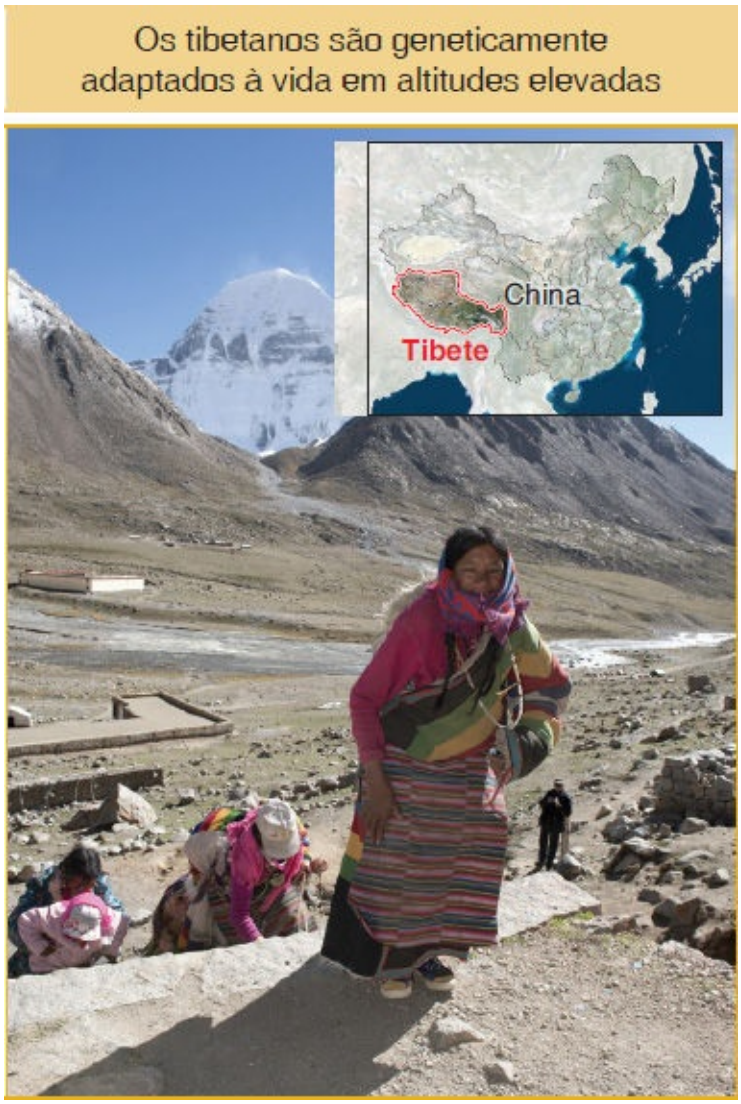


FIGURA 1.21 Uma jovem mulher tibetana. A inserção demonstra a localização do Tibete na Ásia. (Stefan Auth/imagebroker/AGE Fotostock; *inserção*, Planet Observer/UIG/Getty Images.)

Quando a equipe de pesquisadores analisou os seus dados de SNP, os SNP em um gene destacaram-se. O gene é denominado *EPAS1* e alguns SNP nele ocorrem em frequências muito diferentes em tibetanos (87%) e em chineses Han (9%). Seus resultados estão demonstrados na Figura 1.22. Nessa figura, os cromossomos humanos, numerados 1 a 22, estão ao longo do eixo x e uma medida da diferença na frequência da variante do SNP entre tibetanos e chineses está no eixo y. Cada ponto representa um SNP. Os SNP que se encontram acima da linha vermelha horizontal são aqueles em relação aos quais a diferença de frequência

entre tibetanos e chineses Han é tão grande que o gene próximo desses SNP deve ter proporcionado alguma vantagem às pessoas que colonizaram o Platô Tibetano. Os SNP em *EPAS1* encontram-se acima da linha.

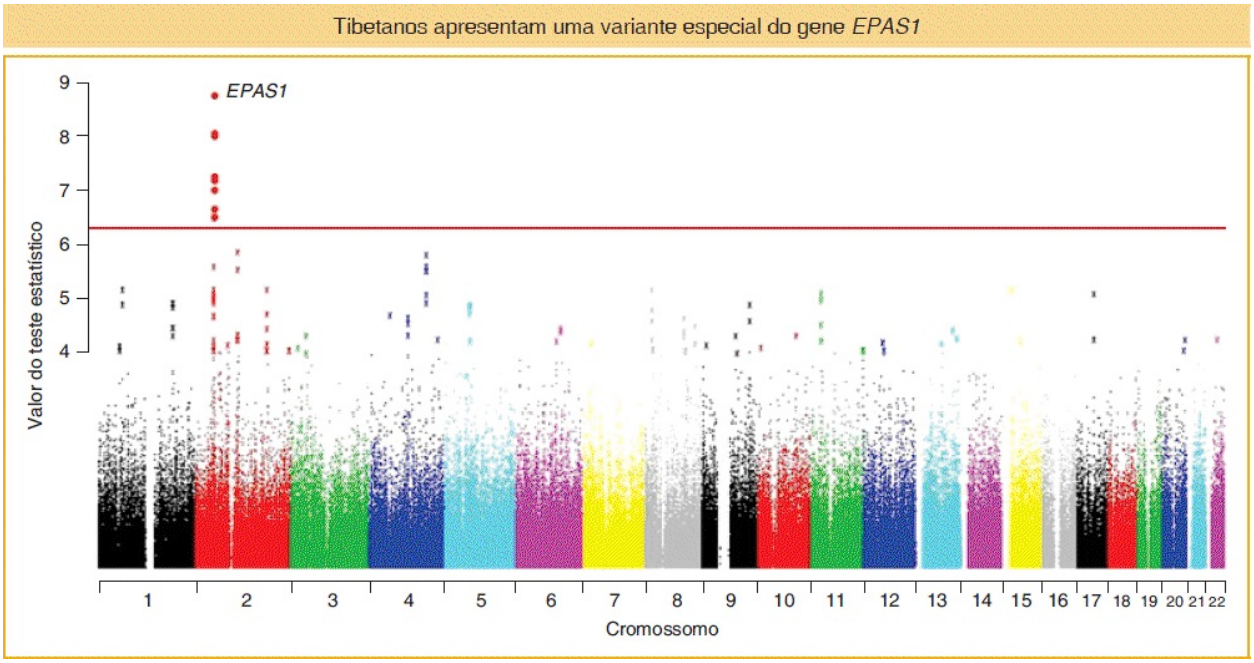


FIGURA 1.22 Os 22 cromossomos humanos são arranjados da esquerda para a direita. O *eixo y* demonstra os resultados de um teste estatístico sobre a possibilidade de existir uma diferença significativa na frequência de SNP entre tibetanos e chineses Han. Cada *pequeno ponto* representa um dos SNP que foram testados. Os SNP *acima da linha vermelha horizontal* são significativamente diferentes. Apenas os SNP no gene *EPAS1* demonstram uma diferença significativa. (C. Beall *et al.*, *Proceedings of the National Academy of Sciences USA*, 107, 25, 2010, 11459-11464, Fig. 1.)

Tais resultados sugerem que os tibetanos apresentam uma variante especial do *EPAS1* que os auxilia na adaptação à vida em altitudes elevadas. Para compreender melhor esse fato, revisemos primeiramente o que se sabe a respeito do *EPAS1*. Esse gene regula a quantidade de eritrócitos que nossos corpos produzem. Regula também a quantidade de eritrócitos em resposta ao nível de oxigênio em nossos tecidos. Quando os níveis de oxigênio em nossos tecidos estão baixos, o *EPAS1* sinaliza o corpo para produzir mais eritrócitos.

Por que o *EPAS1* direciona nossos corpos a produzirem mais eritrócitos quando os níveis de oxigênio em nossos tecidos estão baixos? A resposta do *EPAS1* à baixa concentração de oxigênio pode ser como nossos corpos normalmente

respondem à anemia (muito poucos eritrócitos). Pessoas com baixas contagens de eritrócitos obtêm muito pouco oxigênio em seus tecidos e, assim, o *EPAS1* pode sinalizar o corpo para que fabrique mais eritrócitos para corrigir a anemia. Esse mecanismo pode explicar o motivo pelo qual as pessoas que vivem em baixas altitudes necessitam do gene *EPAS1*.

Agora, pensemos sobre como pessoas de uma baixa altitude responderiam caso se mudassem para uma altitude elevada. Em virtude do ar rarefeito em altas altitudes, seus tecidos obteriam menos oxigênio. Se os seus corpos interpretassem a baixa concentração de oxigênio em virtude do ar rarefeito como um sinal de anemia, o *EPAS1* tentaria corrigir o problema ao sinalizar seus corpos para fabricarem mais eritrócitos. Entretanto, tendo em vista que elas não estão anêmicas e já possuem eritrócitos suficientes, seu sangue se tornaria sobrecarregado com eritrócitos. Uma contagem elevada deles pode causar hipertensão pulmonar e formação de coágulos sanguíneos, condições subjacentes à DCM.

Finalmente, como uma nova variante do *EPAS1* poderia ter ajudado os tibetanos a evitarem a DCM e a se adaptarem às altas altitudes? A resposta para essa questão não é conhecida e atualmente está sendo ativamente investigada, mas existe uma hipótese. Contrariamente aos nativos de planícies, os tibetanos mantêm níveis relativamente normais de eritrócitos em altas elevações e apresentam um risco mais baixo de formação de coágulos sanguíneos e hipertensão pulmonar do que nativos de planícies que se mudam para altas altitudes. Portanto, a versão do *EPAS1* dos tibetanos pode ter deixado de causar a produção excessiva de eritrócitos em altas altitudes, ao mesmo tempo que proporciona outro mecanismo para lidar com o ar rarefeito. A variante tibetana do *EPAS1* ajuda-os a viver em altitudes elevadas sem sofrer de DCM.

Tolerância à lactose. Antes da invenção da agricultura há aproximadamente 10.000 a 12.000 anos, as populações humanas sobreviviam de alimentos colhidos da natureza ao caçar animais selvagens e colher frutas e vegetais selvagens. Naquela época, nenhuma população humana consumia laticínios. O gado ainda precisava ser domesticado e os métodos para ordenha das vacas ainda não

haviam sido inventados. Crianças eram alimentadas com o leite materno, mas quando cresciam, o gene que codifica a enzima *lactase*, a qual possibilita que as crianças realizem a digestão do açúcar do leite (lactose), era desligado. Após o desmame, uma criança nas sociedades pré-agrícolas deixava de necessitar da enzima lactase, de modo que o gene da lactase apresentava um "interruptor" ou elemento regulatório, que o desligava na segunda infância.

Com a origem da agricultura, o gado foi domesticado a partir de auroques selvagens. Os primeiros fazendeiros podem ter mantido o gado primeiramente como uma fonte de carne. Após a invenção da ordenha, o leite oferecia outra fonte de alimento. Mas havia um problema. Embora as crianças nessas sociedades antigas conseguissem digerir o açúcar do leite, os adultos não conseguiam. Os adultos poderiam consumir o leite, mas, tendo em vista que não conseguiam digerir a lactose, apresentavam distenção abdominal, cólicas e diarreia. Os adultos que apresentam esses sintomas em virtude da ingestão de leite são *intolerantes à lactose*. Dessa maneira, tendo em vista que não conseguiam digerir o açúcar do leite, não o estavam utilizando como fonte de nutrição.

Nas sociedades antigas, nas quais os alimentos por vezes poderiam ser escassos, a diferença entre a vida e a morte poderia depender de fazer o melhor uso de todas as fontes de alimentos disponíveis. Ainda assim, tendo em vista que o gene da lactase é desligado em adultos, esses não conseguiam digerir a lactose. Suponha, porém, que uma nova mutação tenha sido inserida na população, mutação essa que tenha possibilitado ao gene da lactase ser expresso em adultos. Adultos com essa nova mutação ou variante então poderiam se beneficiar da ingestão do leite de um modo que adultos com ausência desta variante não poderiam. Tal benefício poderia aumentar as suas chances de sobreviver e ter filhos e, ao longo do tempo, a variante que proporciona a persistência da lactase na vida adulta se tornaria mais comum na população.

O cenário há pouco descrito é o que aparenta ter ocorrido durante a história humana em diversas áreas do mundo nas quais as pessoas mantinham gado (ou camelos) e consumiam o leite desses animais. Isso ocorreu na Europa, no Oriente Médio e na África. Na Europa, algumas pessoas têm uma variante do gene da lactase que apresenta uma "T" em um SNP em particular, enquanto pessoas de

outras regiões do mundo apresentam uma “C” nesse SNP. Recentemente, geneticistas descobriram que a “T” aparenta estar localizada em um elemento regulatório que controla quando o gene da lactase é ligado (Figura 1.23). Pessoas com a variante “T” apresentam expressão persistente do gene da lactase na vida adulta, enquanto pessoas com a variante “C” apresentam seu gene da lactase desligado após a infância. A “T” aparenta possibilitar a ligação de uma proteína regulatória denominada OCT1 próxima ao gene da lactase e, assim, causar a sua expressão em adultos. Outras variantes que apresentam o mesmo efeito aparentam ter surgido de modo independente no Oriente Médio e na Ásia.

Conforme demonstrado na Figura 1.24, no norte da Europa, onde a pecuária e o consumo de laticínios são proeminentes, a persistência da lactase e a variante da lactase “T” que a produz são comuns, enquanto tais características são muito menos comuns no sul da Europa. Geneticistas inferem que os criadores de gado iniciais do norte da Europa que apresentavam a variante “T” se beneficiaram do consumo do leite, o que possibilitou que eles sobrevivessem e produzissem mais descendentes, de modo que essa variante se tornou mais comum na população ao longo do tempo. Atualmente, a variante “T” encontra-se em uma frequência de 90% no norte da Europa. Tendo em vista que o leite não era uma parte tão importante da dieta no sul da Europa, a variante T não oferecia benefício em especial e, assim, permaneceu em uma frequência mais baixa (aproximadamente 10%).

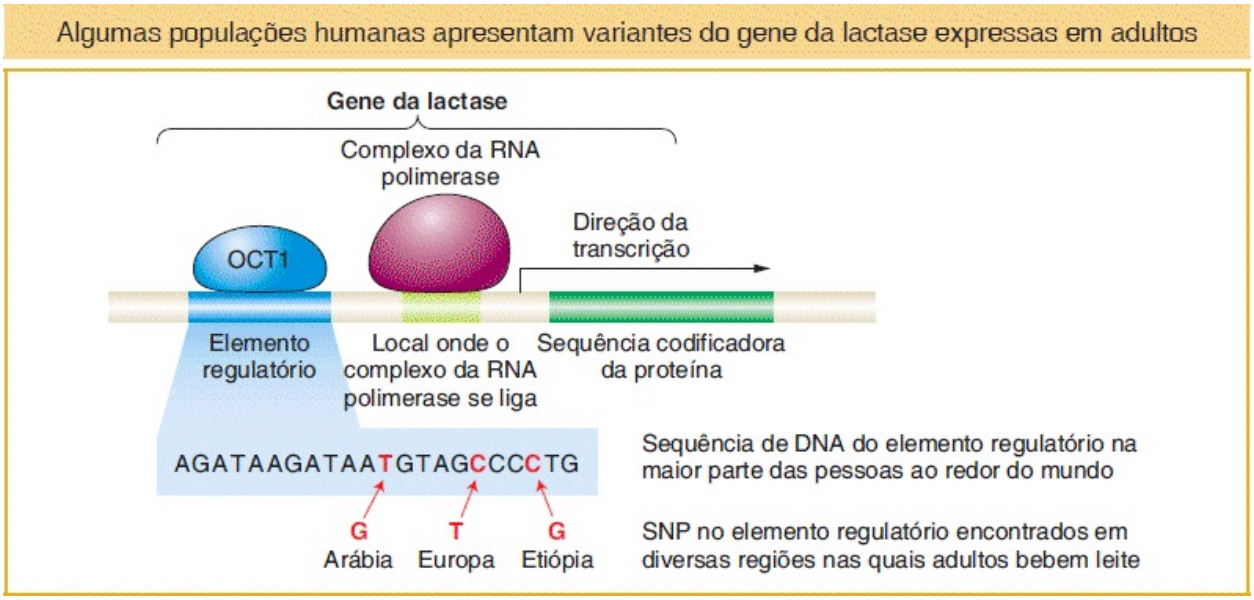


FIGURA 1.23 Diagrama simplificado do gene da lactase demonstrando um elemento regulatório e a região codificadora da proteína. Acredita-se que a OCT1 seja uma proteína reguladora da expressão do gene da lactase. Variantes de SNP no elemento regulatório são encontradas em algumas partes do mundo. Esses SNP estão associados à ligação da OCT1 ao elemento e à expressão do gene da lactase em adultos.

Esses dois exemplos destacam como as populações humanas evoluíram nos tempos recentes em resposta às condições da vida, tais como os alimentos disponíveis e o clima. Nos últimos três capítulos deste livro, você aprenderá a teoria e os métodos utilizados pelos geneticistas para compreender como as populações evoluem em resposta ao seu ambiente. Você aprenderá como os dados de SNP são reunidos, como as frequências das variantes são calculadas e como são realizadas comparações para compreender as forças que influenciaram os tipos de variantes gênicas que ocorrem em diferentes populações. Por meio desse tipo de análise, geneticistas evolutivos aprenderam muito a respeito de como diferentes espécies de plantas, animais, fungos e microrganismos evoluíram e continuam a evoluir em resposta às condições em que vivem.

> CONCEITO-CHAVE Geneticistas evolutivos fornecem as ferramentas para documentar como as variantes gênicas que proporcionam um efeito benéfico podem aumentar em frequência em uma população e tornar os indivíduos na população mais bem-adaptados ao ambiente em que vivem.

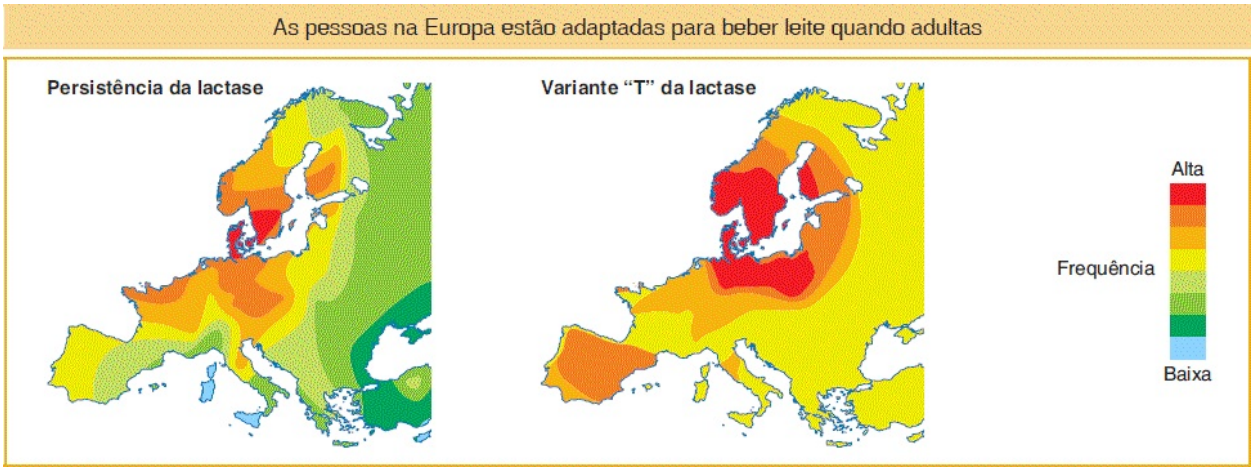


FIGURA 1.24 A. Frequência, na Europa, da persistência da lactase; expressão da enzima lactase em adultos. **B.** Frequência, na Europa, da variante T no gene da lactase, que aparenta controlar a persistência da lactase. (*A.* Adaptada, com autorização, de Y. Itan *et al.*, *BMC Evolutionary Biology* 10, 2010, 36; *B.* Adaptada, com autorização, de A. Beja-Pereira *et al.*, *Nature Genetics* 35, 2003, 311-313.)

RESUMO

Ao iniciar seus estudos sobre genética, imagine-se no meio de uma surpreendente jornada de descobertas. Os últimos 100 anos testemunharam uma revolução extraordinária no conhecimento humano a respeito de como os sistemas biológicos são reunidos e atuam. A genética esteve no epicentro dessa revolução. A análise genética respondeu a muitas questões fundamentais a respeito da transmissão das informações genéticas dentro das famílias e das células, bem como ao longo das eras do período evolutivo. Ainda assim, conforme você aprenderá, o processo de descoberta na genética nunca foi mais dinâmico e o ritmo do crescimento no conhecimento nunca foi maior. As questões não respondidas são abundantes

- Como todos os genes no genoma atuam em conjunto para transformar um ovócito fertilizado em um organismo adulto?
- Como as células conseguem orquestrar continuamente o arranjo incrivelmente complexo das interações gênicas e das reações bioquímicas que ocorrem dentro delas?

- Como as variantes genéticas em centenas ou até mesmo milhares de genes controlam a produção das plantas cultivadas?
- Como a genética pode orientar a prevenção e o tratamento do câncer, do autismo e de outras doenças?
- Como os genes proporcionam aos humanos a capacidade da linguagem e da consciência?

A análise genética ao longo dos próximos 100 anos promete ajudar a responder muitas questões como essas.

TERMOS-CHAVE

adenina (A)
alelos
células somáticas
citosina (C)
códon
complementar
DNA polimerase
dominante
elemento regulatório
expressão gênica
gametas
gene
genética
genômica
guanina (G)
hipótese multifatorial
hipótese um gene–uma enzima
ligase
locus de traço quantitativo (QTL)
mutação de ponto
nuclease

organismo geneticamente modificado (OGM)
organismo-modelo
polimorfismo de nucleotídio único (SNP)
replicação do DNA
RNA mensageiro (mRNA)
sequenciamento do DNA
teoria cromossômica
teoria da mistura
timina (T)
tradução
transcrição
transformação

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Se a variedade parental com flores brancas na Figura 1.3 fosse cruzada com a planta híbrida de primeira geração naquela figura, quais tipos de progênie você esperaria observar e em quais proporções?
2. Na publicação de Mendel de 1866, conforme demonstrado na Figura 1.4, ele relata 705 descendentes com flores roxas (violeta) e 224 descendentes com flores brancas. A proporção que ele obteve é de 3,15:1 para roxas:brancas. Como você acha que ele explicou o fato de que a proporção não é de exatamente de 3:1?
3. Na Figura 1.6, os estudantes apresentam uma de 15 estaturas diferentes e existem duas classes de estatura (1,50 m e 1,52 m) em relação às quais não são observados estudantes. Esse é um total de 17 classes de altura. Se um gene mendeliano único pode explicar apenas duas classes de um traço (como flores roxas ou brancas), quantos genes mendelianos seriam minimamente necessários para explicar a observação de 17 classes de estatura?

4. A Figura 1.7 demonstra uma via simplificada para a síntese da arginina em *Neurospora*. Suponha que você dispõe de uma cepa especial de *Neurospora* que produz citrulina, mas não arginina. Qual(is) gene(s) provavelmente é(são) mutante(s) ou está(ão) ausente(s) na sua cepa especial? Você dispõe de uma segunda cepa de *Neurospora* que não produz citrulina nem arginina, mas produz ornitina. Qual(is) gene(s) é(são) mutante(s) ou está(ão) ausente(s) nessa cepa?

5. Considere a Figura 1.8 A.

a. O que as esferas azuis pequenas representam?

b. O que as pranchas marrons representam?

c. Você concorda com a analogia de que o DNA está estruturado como uma escada?

6. Na Figura 1.8 B, você consegue dizer se a quantidade de ligações de hidrogênio entre adenina e timina é a mesma daquelas entre citosina e guanina? Você acha que uma molécula de DNA com um alto conteúdo de A + T seria mais estável do que uma molécula com um alto conteúdo de G + C?

7. Qual dos três maiores grupos (domínios) da vida na Figura 1.11 não está representado por um organismo-modelo?

8. A Figura 1.15 demonstra o heredograma, ou árvore genealógica, de Louise Benge (indivíduo VI-1), que sofre da doença ACDC por apresentar duas cópias mutantes do gene CD73. Ela tem quatro irmãos (VI-2, VI-3, VI-4 e VI-5), que apresentam a doença, pelo mesmo motivo. Todos os 10 filhos de Louise e seus irmãos apresentam o mesmo número de cópias mutantes do gene CD73 ou esse número pode ser diferente para alguns dos 10 filhos?

PROBLEMAS BÁSICOS

9. Abaixo está a sequência de um filamento único de uma molécula curta de DNA. Em um pedaço de papel, reescreva esta sequência e, em seguida, escreva a sequência do filamento complementar abaixo dela.

GTTCGCGGCCGCGAAC

Comparando os filamentos superior e inferior, o que você observa a respeito da relação entre eles?

10. Mendel estudou uma variedade *alta* de ervilha, com caules que apresentam 20 cm de comprimento, e uma variedade *anã*, com caules que apresentam apenas 12 cm de comprimento.

a. Sob a teoria da mistura, quão longos você espera que os caules dos híbridos de primeira e segunda geração sejam?

b. Sob as regras mendelianas e presumindo que o comprimento do caule seja controlado por um gene único, o que você espera observar nos híbridos de segunda geração se todos os híbridos de primeira geração forem altos?

11. Se uma dupla-hélice de DNA que tem 100 pares de bases de comprimento apresentar 32 adeninas, quantas citosinas, guaninas e timinas ela deve apresentar?

12. Os filamentos complementares do DNA na dupla-hélice são mantidos unidos por meio de pontes de hidrogênio: G ≡ C ou A = T. Essas ligações podem ser quebradas (desnaturadas) em soluções aquosas por meio de aquecimento para produzir dois filamentos únicos de DNA (ver Figura 1.13 A). Como você espera que as quantidades relativas de pares de bases GC *versus* AT em uma dupla-hélice de DNA afetem o calor necessário para a sua desnaturação? Como você espera que o comprimento de uma dupla-hélice de DNA em pares de bases afete o calor necessário para a sua desnaturação?

13. A figura na parte inferior da página demonstra a sequência de DNA de uma parte de um dos cromossomos de um trio (mãe, pai e filho). Você consegue localizar quaisquer novas mutações de ponto no filho que não se encontram em nenhum dos genitores? Em qual genitor surgiu a mutação?

PROBLEMAS DESAFIADORES

14.

a. Existem três nucleotídios em cada códon e cada um deles pode apresentar uma de quatro bases diferentes. Quantos códons únicos possíveis existem ali?

b. Se o DNA apresentasse apenas dois tipos de bases em vez de quatro, quão longos os códons precisariam ser para especificar todos os 20 aminoácidos?

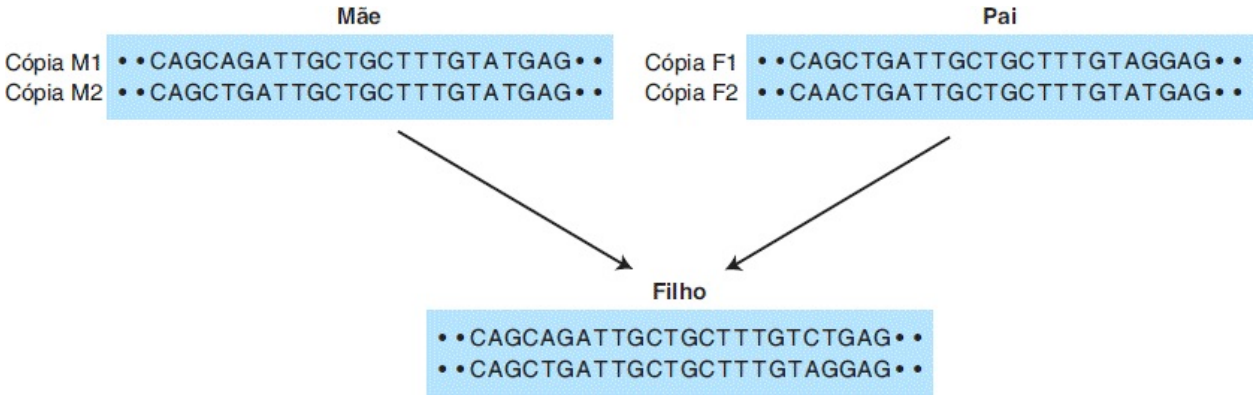
15. Os pais contribuem com mais mutações de ponto novas para seus filhos do que as mães. Você deve saber da biologia geral que as pessoas apresentam cromossomos sexuais — dois cromossomos X em mulheres e um cromossomo X e um Y em homens. Ambos os sexos apresentam os autossomos (A).

a. Em qual tipo de cromossomo (A, X ou Y) você esperara que os genes apresentem a maior quantidade de novas mutações por par de bases ao longo de muitas gerações em uma população? Por quê?

b. Em qual tipo de cromossomo você espera o menor número de novas mutações de par de bases? Por quê?

c. Você consegue calcular o número esperado de novas mutações por par de bases em relação a um gene nos cromossomos X e Y para cada nova mutação em um gene autossômico se a taxa de mutação em homens for o dobro daquela em mulheres?

16. Em relação a homens jovens de 20 anos de idade, ocorreram 150 rodadas de replicação do DNA durante a produção de espermatozoides, em comparação a apenas 23 rodadas para uma mulher de 20 anos de idade. Essa é uma quantidade 6,5 vezes maior de divisões celulares e uma oportunidade proporcionalmente maior para novas mutações de ponto. Ainda assim, em média, homens de 20 anos de idade contribuem com apenas aproximadamente o dobro de novas mutações de ponto para a sua descendência em relação às mulheres. Como você consegue explicar essa discrepância?



17. Na ciência da computação, um *bit* armazena um de dois estados, 0 ou 1. Um *byte* é um grupo de 8 *bits* que apresenta $2^8 = 256$ estados possíveis. Os arquivos de computador moderno com frequência são do tamanho de *megabites* (10^6 *bytes*) ou mesmo *gigabytes* (10^9 *bytes*). O genoma humano tem o tamanho de aproximadamente 3 bilhões de pares de bases. Quantos nucleotídios são necessários para codificar um único *byte*? Quão grande um arquivo de computador deveria ser para armazenar a mesma quantidade de informações de um genoma humano?

18. O genoma humano tem o tamanho de aproximadamente 3 bilhões de pares de bases.

a. Com a utilização de um papel-padrão de 21,6 cm × 27,9 cm com margens de 2,5 cm, um tamanho de fonte 12 e linhas com espaçamento simples, quantas folhas de papel impressas em um lado seriam necessárias para imprimir o genoma humano?

b. Uma resma (500 folhas de papel) apresenta aproximadamente 5 cm de espessura. Quão alta seria a pilha de papel com o genoma humano total?

c. Você desejaria uma mochila, um carrinho de supermercado ou um caminhão para transportar essa pilha?

[1]F. Jacob e J. Monod, *Cold Spring Harbor Quant. Symp. Biol.* 26, 1963, 393.

[2]V. J. Vitzthum, "The home team advantage: Reproduction in women indigenous to high altitude", *Journal of Experimental Biology* 204, 2001, 3141-3150.

CAPÍTULO 2

Herança Monogênica

O monastério do pai da genética, Gregor Mendel. Uma estátua de Mendel encontra-se visível ao fundo. Atualmente, essa parte do monastério é um museu, e os curadores plantaram begônias vermelhas e brancas em uma disposição que representa graficamente o tipo de padrões de herança obtido por Mendel com ervilhas. (Anthony Griffiths.)

TÓPICOS

2.1 Padrões de herança monogênica
2.2 Base cromossômica dos padrões de herança monogênica
2.3 Base molecular dos padrões de herança mendeliana
2.4 Alguns genes descobertos por meio da observação das proporções de segregação
2.5 Padrões de herança monogênica ligada ao sexo
2.6 Análise de heredogramas humanos

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Descobrir um conjunto de genes que afetam uma propriedade biológica específica de interesse, por meio da observação das proporções de herança monogênica de mutantes que afetam aquela propriedade
- Na progênie dos cruzamentos controlados, reconhecer as proporções fenotípicas diagnósticas da herança monogênica (1:1 em haploides, e 3:1, 1:2:1 e 1:1 em diploides)
- Explicar as proporções de herança monogênica em termos do comportamento cromossômico na meiose
- Prever as proporções fenotípicas entre descendentes de cruzamentos de genitores que diferem em um único gene
- Propor hipóteses razoáveis para explicar a dominância e a recessividade de alelos específicos no nível molecular
- Aplicar as regras da herança monogênica para a análise de heredogramas em seres humanos e reconhecer os padrões diagnósticos de condições autossômicas dominantes, autossômicas recessivas, ligadas ao X dominantes e ligadas ao X recessivas
- Calcular o risco de os descendentes herdarem uma condição causada por um alelo mutante em um ou mais ancestrais específicos.

Que tipos de pesquisas os biólogos realizam? Uma área central de pesquisas na biologia de todos os organismos é a tentativa de compreender como um organismo se desenvolve a partir de um ovo fertilizado até um adulto — em outras palavras, o que faz um organismo ser como

ele é. Normalmente, esse objetivo geral é fragmentado no estudo das **propriedades** biológicas individuais, tais como o desenvolvimento da cor das flores da planta ou a locomoção animal ou a absorção de nutrientes, embora os biólogos também estudem algumas áreas gerais, tais como o modo como uma célula funciona. Como os geneticistas analisam as propriedades biológicas? A abordagem genética para a compreensão de qualquer propriedade biológica é encontrar o subconjunto de genes no genoma que influenciam aquela propriedade, um processo por vezes denominado **descoberta dos genes**. Após esses genes terem sido identificados, as suas funções celulares podem ser elucidadas por meio de pesquisas adicionais.

Existem diversos tipos diferentes de abordagens analíticas para a descoberta dos genes, mas um método amplamente utilizado depende da detecção de *padrões de herança monogênica*, tópico deste capítulo.

Tudo na genética, em um ou outro aspecto, tem por base as variantes que podem ser herdadas. A abordagem básica da genética é comparar e contrastar as propriedades de variantes, e a partir dessas comparações fazer deduções a respeito da função genética. Isso é semelhante ao modo como você faria inferências a respeito de como funciona uma máquina não familiar ao alterar a composição ou as posições das partes atuantes ou até mesmo por meio da remoção das partes, cada uma de uma vez. Cada variante representa um "ajuste" da máquina biológica, a partir do qual a sua função pode ser deduzida.

Em genética, o tipo mais comum de qualquer propriedade de um organismo é denominado **tipo selvagem**, o qual é encontrado "no ambiente" ou na natureza. As variantes hereditáveis observadas em um organismo que difere do tipo selvagem são **mutantes**, organismos que apresentam algum tipo anormal de uma propriedade. Como exemplos, o tipo selvagem e alguns mutantes em dois organismos-modelo estão mostrados na Figura 2.1. Os tipos alternativos da propriedade são denominados **fenótipos**. Nessa análise, distinguimos um fenótipo do tipo selvagem e um fenótipo mutante.

Em comparação ao tipo selvagem, os mutantes são raros. Sabemos que eles surgem a partir dos tipos selvagens por meio de um processo denominado **mutação**, que resulta em uma alteração que pode ser herdada no DNA de um

gene. O tipo alterado do gene também é denominado mutação. As mutações nem sempre são prejudiciais para um organismo; por vezes, elas podem ser vantajosas, mas com mais frequência não apresentam efeito observável. Sabe-se bastante a respeito dos mecanismos que causam as mutações (ver Capítulo 16), mas em geral pode-se dizer que elas surgem a partir de erros no processamento celular do DNA.

A maior parte das populações naturais também apresenta **polimorfismos**, definidos como a coexistência de dois ou mais fenótipos de uma propriedade biológica razoavelmente comuns, tais como a ocorrência de ambas as plantas com frutas vermelhas e laranja em uma população de framboesas selvagens. A análise genética pode utilizar (e utiliza) os polimorfismos, mas os polimorfismos apresentam a desvantagem de em geral não envolver a propriedade específica de interesse para o pesquisador. Os mutantes são muito mais úteis, tendo em vista que possibilitam que o pesquisador enfoque qualquer propriedade.

Declarando simplesmente, as etapas gerais da análise funcional por meio da descoberta dos genes são como segue:

1. Reunião de mutantes que afetam a propriedade biológica de interesse.
2. **Cruzamento** (acasalamento) de mutantes com o tipo selvagem para verificar se os seus descendentes mostram proporções de selvagens e mutantes características da herança monogênica.

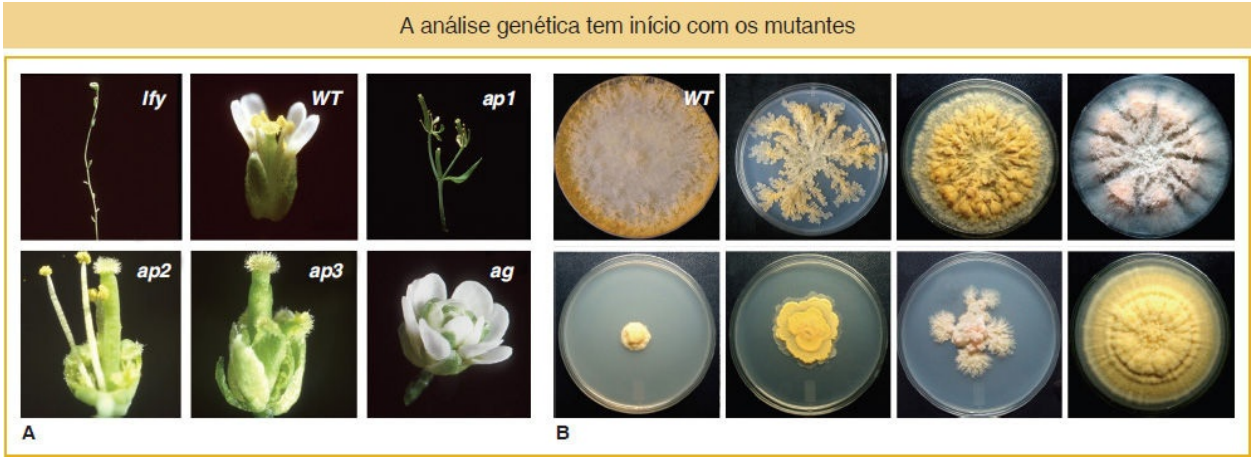


FIGURA 2.1 Estas fotografias mostram a variedade dos fenótipos mutantes típicos daqueles obtidos na dissecção genética das propriedades biológicas. Esses casos são da dissecção do desenvolvimento das flores

na *Arabidopsis thaliana* (A) e do crescimento de hifas em *Neurospora crassa*, um fungo (B). *WT* = Tipo selvagem. (*A.* George Haughn; *B.* Anthony Griffiths/Olivera Gavric.)

3. Dedução das funções do gene no nível molecular.
4. Dedução de como o gene interage com outros genes para produzir a propriedade em questão.

Dessas etapas, apenas 1 e 2 serão abrangidas no presente capítulo.

A descoberta dos genes tem início com uma "caçada" para reunir mutantes nos quais a função biológica em investigação está alterada ou destruída. Embora os mutantes sejam individualmente raros, existem modos de intensificar a sua recuperação. Um método amplamente utilizado é tratar o organismo com radiação ou substâncias químicas que aumentam a taxa de mutação. Após o tratamento, o modo mais direto de identificar os mutantes é rastrear visualmente uma quantidade muito grande de indivíduos, procurando por uma ocorrência aleatória de mutantes naquela população. Além disso, diversos métodos de seleção podem ser planejados para o enriquecimento em relação aos tipos procurados.

Armados com um conjunto de mutantes que afetam a propriedade de interesse, espera-se que cada mutante represente uma lesão em um de um conjunto de genes que controlam a propriedade. Portanto, a esperança é que uma via genética ou uma rede razoavelmente completa esteja representada. Entretanto, nem todos os mutantes são causados por lesões monogênicas (alguns têm etiologia muito mais complexa), de modo que primeiramente cada mutante deve ser testado para verificar se de fato ele é causado por uma mutação monogênica.

O teste em relação à herança monogênica ocorre por meio do cruzamento dos indivíduos que apresentam a propriedade mutante com o tipo selvagem e em seguida com a análise da primeira e da segunda gerações dos descendentes. Como um exemplo, uma planta mutante com flores brancas é cruzada com o tipo selvagem que apresenta flores vermelhas. A progênie desse cruzamento é analisada e, em seguida, intercruzada para produzir uma segunda geração de descendentes. Em cada geração, as proporções diagnósticas de plantas com flores vermelhas em relação àquelas com flores brancas revelam se um gene único controla a cor da flor. Em caso afirmativo, por inferência, o tipo selvagem seria

codificado pela forma selvagem do gene e o mutante seria codificado por um tipo do mesmo gene no qual um evento de mutação tenha de algum modo alterado a sequência do DNA. Outras mutações que afetam a cor da flor (talvez malva, manchada, listrada, e assim por diante) são analisadas do mesmo modo, em geral resultando em um conjunto definido de "genes de cor da flor". A utilização de mutantes desse modo por vezes é denominada **dissecção genética**, tendo em vista que a propriedade biológica em questão (nesse caso, a cor da flor) é selecionada e separada para revelar o seu programa genético subjacente, não com um bisturi, mas com mutantes. Cada mutante possivelmente identifica um gene em separado que afeta aquela propriedade.

Após um conjunto de genes-chave ter sido definido desse modo, diversos métodos moleculares diferentes podem ser utilizados para estabelecer as funções de cada um dos genes. Esses métodos serão abordados nos capítulos posteriores. Portanto, a genética tem sido utilizada para definir o conjunto de funções gênicas que interagem para produzir a propriedade que denominamos cor da flor (neste exemplo).

Esse tipo de abordagem para a descoberta dos genes por vezes é denominado **genética direta**, uma estratégia para a compreensão da função biológica que tem início com mutantes de gene único aleatórios e que termina com a sua sequência de DNA e a sua função bioquímica. (Veremos a **genética reversa** em atuação nos capítulos posteriores. Em resumo, ela tem início com a análise genômica no nível do DNA para identificar um conjunto de genes candidatos que codificam a propriedade biológica de interesse, em seguida induz mutantes direcionados especificamente para aqueles genes, e em seguida examina os fenótipos mutantes para verificar se de fato eles afetam a propriedade em estudo.)

> CONCEITO-CHAVE A abordagem genética para a compreensão de uma propriedade biológica é descobrir os genes que a controlam. Uma abordagem para a descoberta dos genes é isolar os mutantes e verificar cada um em relação aos padrões de herança monogênica (proporções específicas da expressão normal e mutante da propriedade nos descendentes).

A descoberta dos genes é importante não apenas em organismos experimentais, como também em estudos aplicados. Uma área crucial é a agricultura, na qual a descoberta dos genes pode ser utilizada para compreender uma propriedade comercial desejável de um organismo, tal como o seu conteúdo proteico. A genética humana é outra área importante: saber quais funções gênicas estão envolvidas em uma doença ou uma condição específica são informações úteis para desenvolver terapias.

As regras da herança monogênica foram elucidadas originalmente na década de 1860 pelo monge Gregor Mendel, que vivia em um monastério na cidade de Brno, atualmente parte da República Tcheca. A análise de Mendel é o protótipo da abordagem experimental para a descoberta de gene único, ainda utilizada atualmente. De fato, Mendel foi o primeiro a descobrir um gene! Mendel não sabia o que os genes eram, como influenciavam as propriedades biológicas ou como eram herdados no nível celular. Atualmente, sabemos que os genes atuam por meio de proteínas, um tópico ao qual retornaremos nos capítulos posteriores. Também sabemos que os padrões de herança monogênica são produzidos em virtude de os genes serem parte dos cromossomos e de esses serem repartidos de modo muito preciso entre as gerações, conforme veremos posteriormente no capítulo.

2.1 Padrões de herança monogênica

Relembre que a primeira etapa na dissecção genética é obter as variantes que diferem na propriedade em investigação. Supondo que obtivemos uma coleção de mutantes relevantes, a próxima questão é se cada uma das mutações é herdada como um gene único.

Experimentos pioneiros de Mendel

A primeira análise da herança monogênica como uma via para a descoberta dos genes foi realizada por Gregor Mendel. É dele a análise que seguiremos como um exemplo. Mendel escolheu a ervilha de jardim, *Pisum sativum*, como seu organismo de pesquisa. A escolha do organismo para qualquer pesquisa biológica

é crucial, e a escolha de Mendel comprovou ser uma boa escolha, tendo em vista que ervilhas são de fácil plantio e cultivo. Entretanto, observe que Mendel não investigou mutantes de ervilhas; em vez disso, fez uso de mutantes que haviam sido encontrados por outras pessoas e que haviam sido utilizados na horticultura. Além disso, o trabalho de Mendel difere da maior parte das pesquisas genéticas realizadas atualmente, no sentido em que não foi uma dissecção genética; ele não estava interessado nas propriedades das próprias ervilhas, mas sim no modo como as unidades hereditárias que influenciavam aquelas propriedades eram herdadas de geração para geração. Contudo, as leis da herança deduzidas por Mendel são exatamente aquelas que utilizamos atualmente na genética moderna na identificação dos padrões monogênicos de herança.

Mendel optou por investigar a herança de sete propriedades da espécie de ervilha escolhida: cor da ervilha, formato da ervilha, cor da vagem, formato da vagem, cor da flor, altura da planta e posição do broto florescente. Em genética, os termos **característica** e **traço** são utilizados de modo mais ou menos sinônimo; eles significam aproximadamente "propriedade". Para cada uma dessas sete características, Mendel obteve de seu fornecedor de horticultura duas linhagens que apresentavam fenótipos distintos e contrastantes. Esses fenótipos contrastantes estão ilustrados na Figura 2.2. Seus resultados foram substancialmente os mesmos em relação a cada traço e, assim, podemos utilizar um traço, a cor da semente da ervilha, como uma ilustração. Todas as linhagens utilizadas por Mendel eram **linhagens puras**, o que significa que, em relação ao fenótipo em questão, toda a prole produzida por meio de cruzamentos entre os membros daquela linhagem era idêntica. Por exemplo, dentro da linhagem com sementes amarelas, toda a progênie de qualquer cruzamento era de sementes amarelas.

A análise de Mendel da hereditariedade em ervilhas fez uso extensivo de cruzamentos. Para realizar um cruzamento em plantas como a ervilha, o pólen simplesmente é transferido das anteras de uma planta para o estigma de outra. Um tipo especial de cruzamento é a **autopolinização**, que é realizada ao permitir que o pólen de uma flor seja depositado sobre o seu próprio estigma. O cruzamento e a autopolinização estão ilustrados na Figura 2.3. O primeiro cruzamento realizado

por Mendel foi de plantas das linhagens com sementes amarelas com plantas das linhagens com sementes verdes. Em seu programa de cruzamento geral, essas linhagens constituíram a **geração parental**, abreviada **P**. Na *Pisum sativum*, a cor da semente (a ervilha) é determinada pela constituição genética da própria semente; portanto, as ervilhas que resultam de um cruzamento são efetivamente progênie e podem ser convenientemente classificadas em relação ao fenótipo sem a necessidade de seu cultivo até plantas. Observou-se que a progênie de ervilhas do cruzamento entre as diferentes linhagens puras eram todas amarelas, não importava qual genitor (amarelo ou verde) fosse utilizado como masculino ou feminino. Essa geração de progênie é denominada **primeira geração filial**, ou $\mathbf{F_1}$. A palavra *filial* advém das palavras em latim *filia* (filha) e *filius* (filho). Portanto, os resultados desses dois cruzamentos recíprocos foram como segue, em que × representa um cruzamento:

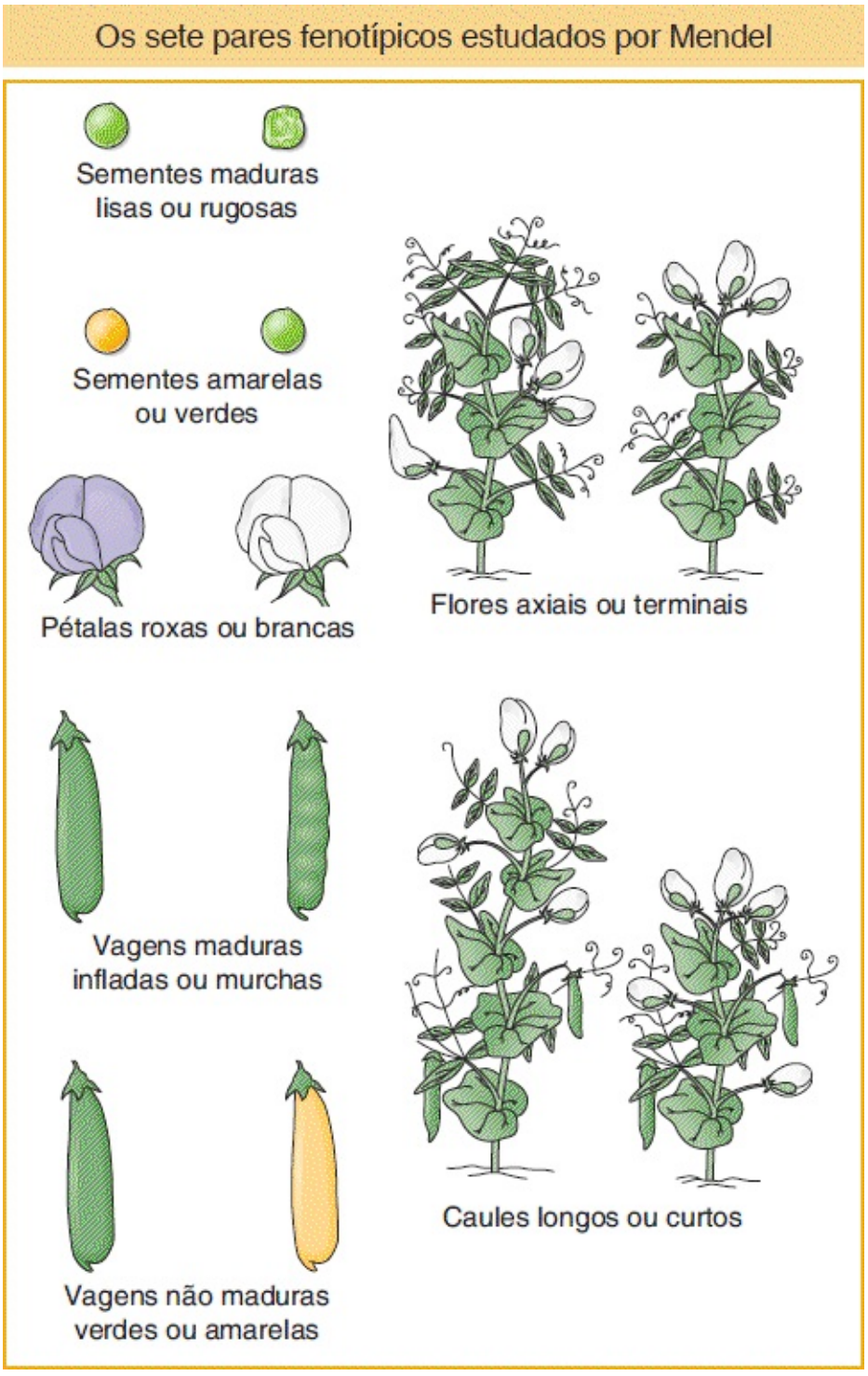


FIGURA 2.2 Para cada característica, Mendel estudou dois fenótipos contrastantes.

Fêmea da linhagem amarela × Macho da linhagem verde → Todas as ervilhas da F_1 amarelas

Fêmea da linhagem verde × Macho da linhagem amarela → Todas as ervilhas da F_1 amarelas

Os resultados observados nos descendentes de ambos os cruzamentos recíprocos foram os mesmos e, assim, iremos tratá-los como um cruzamento. Mendel cultivou as ervilhas da F_1 até se tornarem plantas, e autopolinizou essas plantas para obter a **segunda geração filial**, ou **F_2**. A F_2 foi composta por 6.022 ervilhas amarelas e 2.011 ervilhas verdes. Em resumo:

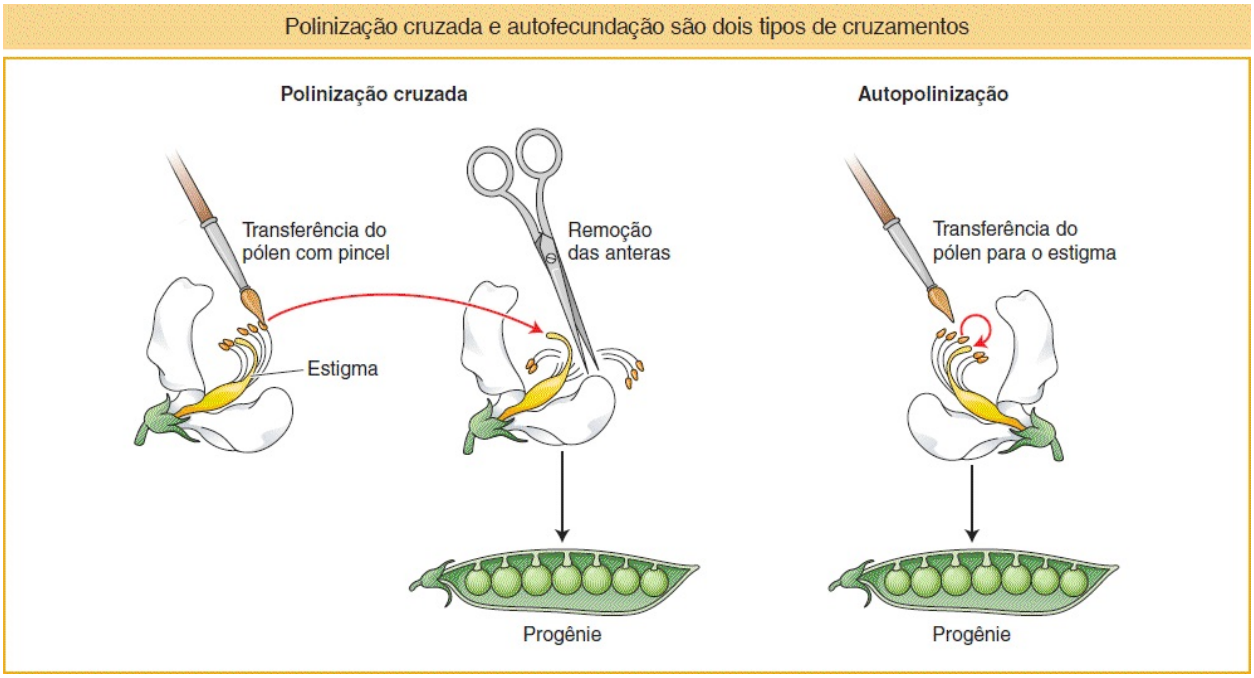


FIGURA 2.3 Em um cruzamento de uma ervilha (*esquerda*), o pólen das anteras de uma planta é transferido para o estigma de outra. Na autopolinização (*direita*), o pólen é transferido das anteras para o estigma da mesma planta.

F_1 amarela × F_1 amarela → F_2 composta por 6.022 amarelas 2.001 verdes Total 8.023

Mendel observou que esse resultado estava muito próximo de uma proporção matemática de três quartos (75%) de amarelas e um quarto (25%) de verdes. Um cálculo simples nos demonstra que 6.022/8.023 = 0,751 ou 75,1%, e que 2.001/8.023 = 0,249 ou 24,9%. Portanto, havia uma proporção de 3:1 de

amarelas e verdes. Curiosamente, o fenótipo verde, que havia desaparecido na F_1, havia reaparecido em um quarto dos indivíduos da F_2, demonstrando que os determinantes genéticos em relação ao verde realmente existiam na F_1 amarela, embora não expressos.

Para investigar adicionalmente a natureza das plantas da F_2, Mendel autopolinizou as plantas cultivadas a partir das sementes da F_2. Ele observou três tipos de resultados diferentes. As plantas cultivadas a partir das sementes verdes da F_2, quando autopolinizadas, originavam apenas ervilhas verdes. Entretanto, a autopolinização das plantas cultivadas a partir das sementes amarelas da F_2 resultou em dois tipos: um terço delas era de sementes amarelas puras, mas dois terços delas deram progênie mista, sendo três quartos de sementes amarelas e um quarto de sementes verdes, assim como as plantas da F_1. Em resumo:

$\frac{1}{4}$ da F_2 era verde, que, quando autopolinizada, forneceu todas verdes

$\frac{3}{4}$ da F_2 eram amarelas; dessas,

$\frac{1}{3}$, quando autopolinizado, originou todas amarelas

$\frac{2}{3}$, quando autopolinizados, originaram $\frac{3}{4}$ amarelas e

$\frac{1}{4}$ verde

Portanto, observada de outro modo, a F_2 foi composta por:

$\frac{1}{4}$ verde puras

$\frac{1}{4}$ amarelo puras

$\frac{1}{2}$ de amarelas semelhantes às da F_1 (progênie mista)

Assim, a proporção de 3:1 em um nível mais fundamental é uma proporção de 1:2:1.

Mendel realizou outro cruzamento informativo entre as plantas com sementes amarelas da F_1 e qualquer planta com sementes verdes. Nesse cruzamento, a progênie demonstrou as proporções de metade amarelas e metade verdes. Em resumo:

$$\text{Amarela da } F_1 \times \text{Verde} \rightarrow \begin{matrix} \frac{1}{2} \text{ amarelo} \\ \frac{1}{2} \text{ verde} \end{matrix}$$

Esses dois tipos de cruzamentos, a autopolinização de F_1 e o cruzamento da F_1 com qualquer planta com sementes verdes, forneceram progênies amarelas e verdes, mas em diferentes proporções. Essas duas proporções estão representadas na Figura 2.4. Observe que as proporções são observadas apenas quando são combinadas as ervilhas de diversas vagens.

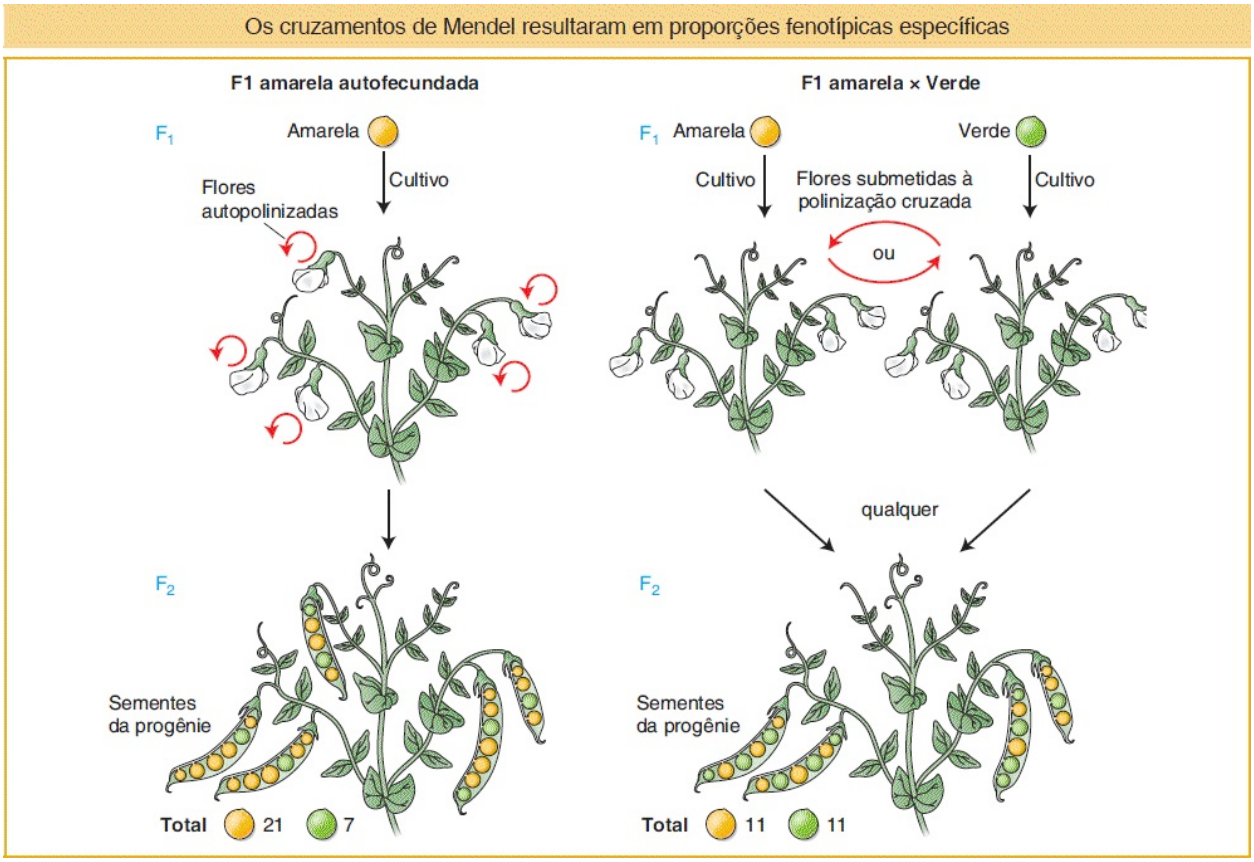


FIGURA 2.4 Mendel obteve uma proporção fenotípica de 3:1 em sua autopolinização da F_1 (*esquerda*) e uma proporção fenotípica de 1:1 em seu cruzamento da F_1 amarela com verde (*direita*). Os tamanhos de amostra são arbitrários.

As proporções de 3:1 e 1:1 observadas em relação à cor da ervilha também foram observadas para cruzamentos comparáveis em relação às outras seis características que Mendel estudou. Os valores reais para as proporções de 3:1

em relação àqueles traços são mostrados na Tabela 2.1.

Lei de Mendel da segregação igual

Inicialmente, o significado dessas proporções matemáticas precisas e repetíveis não foi evidente para Mendel, mas ele foi capaz de planejar um modelo brilhante que não apenas explicou todos os resultados, como também representou o nascimento histórico da ciência da genética. O modelo de Mendel para o exemplo da cor da ervilha, traduzido em termos modernos, foi como segue:

1. Um fator hereditário, denominado **gene**, é necessário para a produção da cor da ervilha.
2. Cada planta apresenta um par desse tipo de gene.
3. O gene se apresenta em duas formas, denominadas **alelos**. Se o gene for foneticamente denominado um gene "ípsilon", então os dois alelos podem ser representados por *Y* (que faz referência ao fenótipo amarelo) e *y* (que faz referência ao fenótipo verde).

Tabela 2.1 Resultados de todos os cruzamentos de Mendel nos quais os genitores diferiam em uma característica.

Fenótipos parentais	**F_1**	**F_2**	**Proporção da F_2**
1. Sementes lisas × Rugosas	Todas lisas	5.474 lisas; 1.850 rugosas	2,96:1
2. Sementes amarelas × Verdes	Todas amarelas	6.022 amarelas; 2.011 verdes	3,01:1
3. Pétalas roxas × Brancas	Todas roxas	705 roxas; 224 brancas	3,15:1

4. Vagens infladas × Murchas	Todas infladas	882 infladas; 299 murchas	2,95:1
5. Vagens verdes × Amarelas	Todas verdes	428 verdes; 152 amarelas	2,82:1
6. Flores axiais × Terminais	Todas axiais	651 axiais; 207 terminais	3,14:1
7. Caules longos × Curtos	Todos longos	787 longos; 277 curtos	2,84:1

4. Uma planta pode ser *Y*/*Y*, *y*/*y*, ou *Y*/*y*. A barra demonstra que os alelos são um par.

5. Na planta *Y*/*y*, o alelo *Y* domina e, assim, o fenótipo será amarelo. Portanto, o fenótipo da planta *Y*/*y* define o alelo *Y* como **dominante** e o alelo *y* como **recessivo**.

6. Na meiose, os membros de um par de genes separam-se igualmente dentro das células que se tornam ovócitos e espermatozoides, os *gametas*. Essa separação igual se tornou conhecida como **primeira lei de Mendel**, ou **lei de segregação igual**. Portanto, um único gameta contém apenas um membro do par de genes.

7. Na fertilização, os gametas se fundem aleatoriamente, independentemente de qual alelos eles contêm.

Aqui, introduzimos uma parte da terminologia. Um ovócito fertilizado, a primeira célula que se desenvolve em um indivíduo da progênie, é denominado **zigoto**. Uma planta com um par de alelos idênticos é denominada **homozigota** (substantivo homozigose), e uma planta na qual os alelos do par são diferentes é denominada **heterozigota** (substantivo heterozigose). Por vezes, um heterozigoto

em relação a um gene é denominado **mono-híbrido**. Um indivíduo pode ser classificado como **homozigoto dominante** (tal como *Y*/*Y*), **heterozigoto** (*Y*/*y*), ou **homozigoto recessivo** (*y*/*y*). Em geral, em genética as combinações alélicas subjacentes aos fenótipos são denominadas **genótipos**. Portanto, *Y*/*Y, Y*/*y* e *y*/*y* são todos genótipos.

A Figura 2.5 demonstra como os postulados de Mendel explicam as proporções das progênies ilustradas na Figura 2.4. As linhagens puras são homozigotas, sejam *Y*/*Y* ou *y*/*y*. Portanto, cada linhagem produz apenas gametas *Y* ou apenas gametas *y*, originando só homozigotos. Quando cruzadas entre si, as linhagens *Y*/*Y* e *y*/*y* produzem uma geração F_1 composta por todos indivíduos heterozigotos (*Y*/*y*). Tendo em vista que *Y* é dominante, todos os indivíduos da F_1 são de fenótipo amarelo. Pode-se pensar na autopolinização dos indivíduos da F_1 como um cruzamento do tipo *Y*/*y* × *Y*/*y*, que por vezes é denominado um **cruzamento mono-híbrido**. A segregação igual dos alelos *Y* e *y* na F_1 heterozigota resulta em gametas do sexo masculino e feminino, metade dos quais são *Y* e metade dos quais são *y*. Os gametas masculinos e femininos se fundem aleatoriamente na fertilização, com os resultados demonstrados na grade na Figura 2.5. A composição da F_2 é de três quartos de sementes amarelas e um quarto de verdes, uma proporção de 3:1. A proporção de um quarto das sementes F_2 é verde puro, conforme esperado do genótipo *y*/*y*. Entretanto, as sementes amarelas da F_2 (que totalizam três quartos) são de dois genótipos: dois terços delas são claramente heterozigotos *Y*/*y*, e um terço é homozigoto dominante *Y*/*Y*. Portanto, observamos que, na base da proporção fenotípica de 3:1 na F_2, encontra-se uma proporção genotípica de 1:2:1:

$$\left.\begin{array}{l}\frac{1}{4}\text{ amarelo } Y/Y\\ \frac{2}{4}\text{ amarelos } Y/y\end{array}\right\}\ \frac{3}{4}\text{ amarelos } (Y/-)$$
$$\frac{1}{4}\text{ verde } y/y$$

A ilustração geral de um indivíduo que expressa o alelo dominante é *Y*/—; o travessão representa uma abertura que pode ser preenchida por outro *Y* ou por um *y*. Observe que a segregação igual é detectável apenas na meiose de um

heterozigoto. Portanto, *Y*/*y* produz metade dos gametas *Y* e metade dos gametas *y*. Embora a segregação igual esteja ocorrendo também em homozigotos, nem a segregação $\frac{1}{2}$ *Y*: $\frac{1}{2}$ *Y*, nem a segregação $\frac{1}{2}$ *y*: $\frac{1}{2}$ *y* é significativa ou detectável no nível genético.

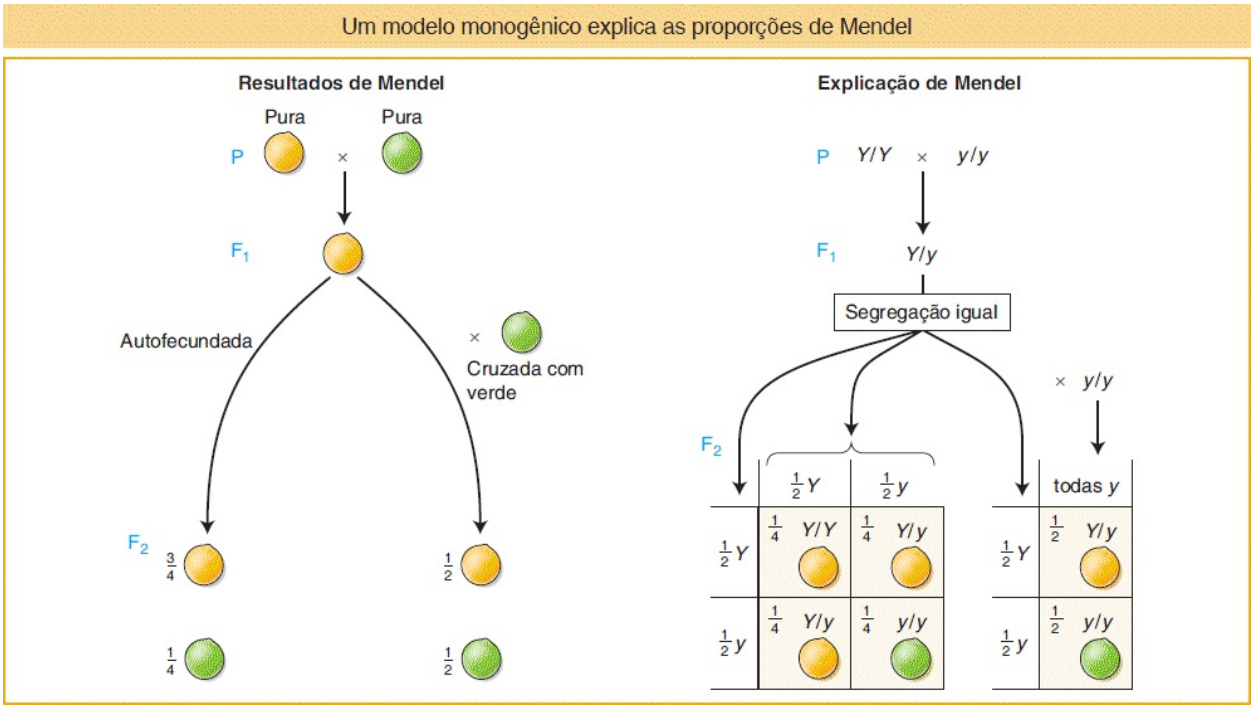


FIGURA 2.5 Os resultados de Mendel (*esquerda*) são explicados por um modelo monogênico (*direita*), que postula a segregação igual dos membros de um par de genes nos gametas.

Agora também podemos explicar os resultados do cruzamento entre as plantas cultivadas a partir de sementes amarelas da F_1 (*Y*/*y*) e as plantas cultivadas a partir de sementes verdes (*y*/*y*). Nesse caso, a segregação igual nas heterozigotas amarelas da F_1 fornece gametas com uma proporção de $\frac{1}{2}$ *Y*: $\frac{1}{2}$ *y*. Entretanto, o genitor *y*/*y* somente pode produzir gametas *y*. Assim, o fenótipo da progênie depende apenas de qual alelo elas herdam do genitor *Y*/*y*. Portanto, a proporção *gamética* de $\frac{1}{2}$ *Y*: $\frac{1}{2}$ *y* do heterozigoto é convertida em uma proporção *genotípica* de $\frac{1}{2}$ *Y*/*y*: $\frac{1}{2}$ *y*/*y*, que corresponde a uma proporção *fenotípica* de 1:1 de plantas com sementes amarelas e plantas com sementes verdes. Isso está ilustrado no painel à direita da Figura 2.5.

Observe que, ao definir os pares de alelos subjacentes aos seus fenótipos,

Mendel havia identificado um gene que afeta radicalmente a cor da ervilha. Essa identificação não era o seu interesse primário, mas podemos observar como o achado dos padrões de herança monogênica é um processo de descoberta de genes, que identifica os genes individuais que influenciam uma propriedade biológica.

> CONCEITO-CHAVE Todas as proporções genéticas de 1:1, 3:1 e 1:2:1 são diagnósticas da herança monogênica e têm por base a segregação igual em um heterozigoto.

A pesquisa de Mendel em meados do século 19 não foi notada pela comunidade científica internacional até que observações semelhantes foram publicadas de modo independente por diversos outros pesquisadores no século 20. Logo depois, pesquisas em muitas espécies de plantas, animais, fungos e algas demonstraram que a lei da segregação igual de Mendel era aplicável a todos os eucariotos sexuados e, em todos os casos, tinha por base as segregações cromossômicas que ocorrem na meiose, um tópico ao qual nos voltaremos na próxima seção.

2.2 Base cromossômica dos padrões de herança monogênica

A consideração de Mendel sobre a segregação igual era que os membros de um par de genes segregavam igualmente *na formação dos gametas*. Ele não tinha conhecimento a respeito dos eventos subcelulares que ocorrem quando as células se dividem na formação dos gametas. Atualmente, compreendemos que os pares de genes estão localizados em pares de cromossomos e que são os membros de um par de cromossomos que de fato se segregam, carreando os genes com eles. Os membros de um par de genes são segregados como uma consequência inevitável.

Herança monogênica em diploides

Quando as células se dividem, assim também precisam se dividir o núcleo e seus principais componentes, os cromossomos. Para compreender a segregação gênica, primeiramente é preciso compreender e contrastar os dois tipos de divisões nucleares que ocorrem em células eucarióticas. Quando as células somáticas (corporais) se dividem para aumentar sua quantidade, a divisão nuclear que as acompanha é denominada **mitose**, um estádio programado de todos os ciclos de divisão celular eucariótica (Figura 2.6). A mitose pode ocorrer em células diploides ou haploides. Como resultado, uma célula genitora se torna duas células geneticamente idênticas. Portanto,

$$\text{Qualquer } 2n \rightarrow 2n + 2n$$
$$\text{ou } n \rightarrow n + n$$

Esse "truque" de constância é conquistado quando cada cromossomo se replica para produzir duas cópias idênticas dele próprio, com a concomitante replicação do DNA. As duas cópias idênticas, que com frequência são discerníveis visualmente, são denominadas **cromátides**-irmãs. Em seguida, cada cópia é puxada em direção às extremidades opostas da célula. Quando a célula se divide, cada célula-filha apresenta o mesmo conjunto cromossômico que a sua genitora.

Além disso, a maior parte dos eucariotos apresenta um ciclo sexuado e, nesses organismos, células diploides especializadas, denominadas **meiócitos**, são separadas e se dividem para produzir células sexuais, tais como espermatozoides e ovócitos em plantas e animais, ou esporos sexuais em fungos ou algas. Ocorrem duas divisões celulares sequenciais, e as duas divisões nucleares que as acompanham são denominadas **meiose**. Tendo em vista que ocorrem duas divisões, quatro células são produzidas a partir de cada célula genitora. A meiose ocorre apenas em células diploides, e os gametas resultantes (espermatozoides e ovócitos em animais e plantas) são haploides. Portanto, o resultado líquido da meiose é:

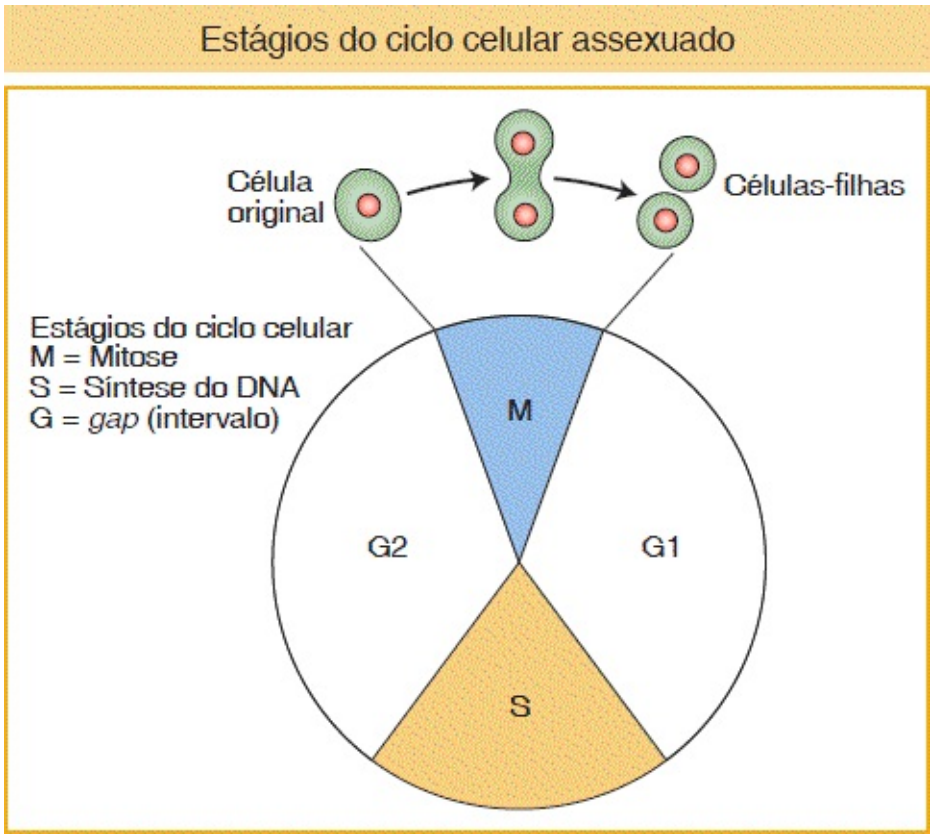


FIGURA 2.6.

$$2n \rightarrow n + n + n + n$$

Essa divisão geral pela metade da quantidade de cromossomos durante a meiose é conquistada por meio de uma replicação e duas divisões. Assim como com a mitose, cada cromossomo é replicado uma vez, mas na meiose os cromossomos replicados (cromátides-irmãs) permanecem unidos. Um de cada um dos pares de cromossomos replicados é puxado em direção às extremidades opostas da célula, e ocorre a divisão. Na segunda divisão, as cromátides-irmãs se separam e são puxadas em direção às extremidades opostas da célula.

A localização dos meiócitos nos ciclos de vida de animais, plantas e fungos está demonstrada na Figura 2.7.

As características *genéticas* básicas da mitose e da meiose estão resumidas na Figura 2.8. Para facilitar a comparação, ambos os processos estão demonstrados em uma célula diploide. Observe, novamente, que a mitose ocorre em uma

divisão celular, e as duas células-“filhas” resultantes apresentam o mesmo conteúdo genômico que a célula-“mãe” (genitora). O primeiro processo-chave a ser observado é uma replicação cromossômica pré-mitótica. No nível do DNA, esse estágio é a fase de síntese, ou S (ver Figura 2.6), na qual o DNA é replicado. A replicação produz pares de cromátides-irmãs idênticas, que se tornam visíveis no início da mitose. Quando uma célula se divide, cada membro de um par de cromátides-irmãs é puxado para dentro de uma célula-filha, onde assume o papel de um cromossomo pleno. Portanto, cada célula-filha apresenta o mesmo conteúdo cromossômico da célula original.

Antes da meiose, assim como na mitose, ocorre a replicação cromossômica para formar cromátides-irmãs, que se tornam visíveis na meiose. O centrômero aparentemente não se divide nesse estádio, mas se divide na mitose. Também contrariamente à mitose, os pares homólogos de cromátides-irmãs agora se unem para formar um feixe de quatro cromátides homólogas. Essa reunião dos pares homólogos é denominada *sinapse*, e se forma em virtude das propriedades de uma estrutura macromolecular denominada complexo sinaptonêmico (SC), a qual corre pelo centro do par. Os cromossomos-irmãos replicados em conjunto são denominados **díade** (da palavra grega “dois”). A unidade que compreende o par de díades em sinapse é denominada **bivalente**. As quatro cromátides que compõem um bivalente são denominadas **tétrade** (da palavra grega “quatro”), para indicar que existem quatro unidades homólogas no feixe.

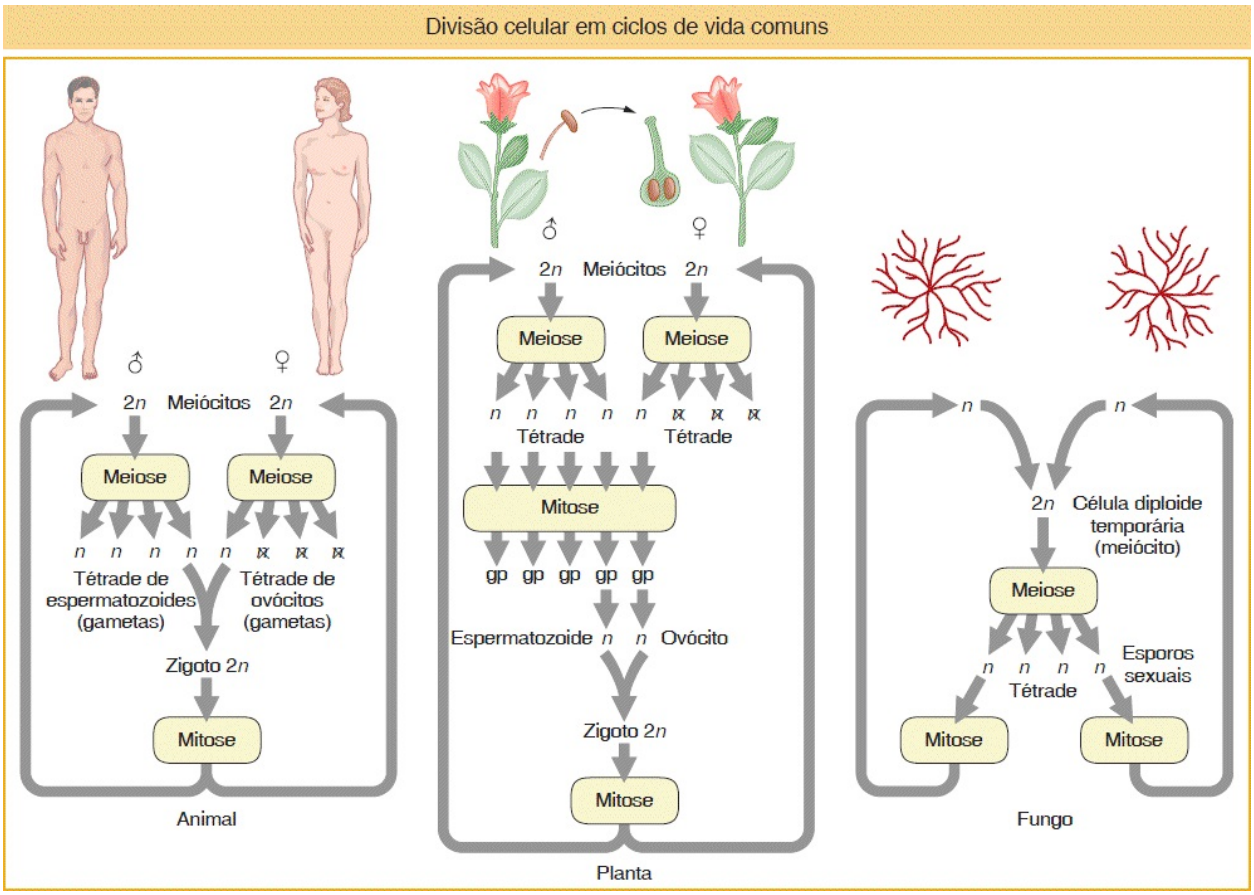
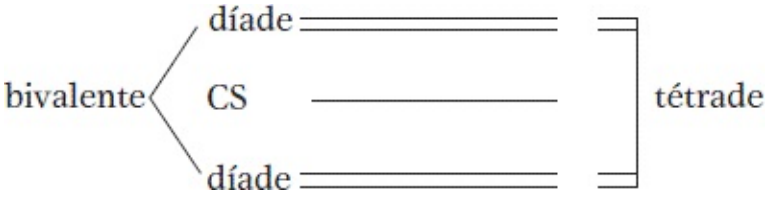


FIGURA 2.7 Os ciclos de vida de humanos, plantas e fungos, demonstrando os pontos nos quais ocorrem a mitose e a meiose. Observe que em humanos do sexo feminino e em muitas plantas, três células da tétrade meiótica são abortadas. A abreviação *n* indica uma célula haploide, 2*n*, uma célula diploide; gp faz referência a gametófito, o nome da pequena estrutura composta por células haploides que produzirá gametas. Em muitas plantas, tais como o milho, um núcleo do gametófito masculino se funde com dois núcleos do gametófito feminino, dando origem a uma célula triploide (3*n*), que em seguida replica para formar o endosperma, um tecido nutritivo que circunda o embrião (que é derivado do zigoto 2*n*).



(Uma observação entre parênteses. O processo de *crossing over* ocorre nesse estágio de tétrade. O *crossing over* altera as combinações de alelos de diversos genes diferentes, mas não afeta diretamente os padrões de herança monogênica; portanto, adiaremos a sua abrangência detalhada até o Capítulo 4. Para o

momento, vale observar que, à parte da sua função de combinação de alelos, o *crossing over* sabidamente também é um evento crucial, que é essencial para a adequada segregação dos cromossomos na primeira divisão meiótica.)

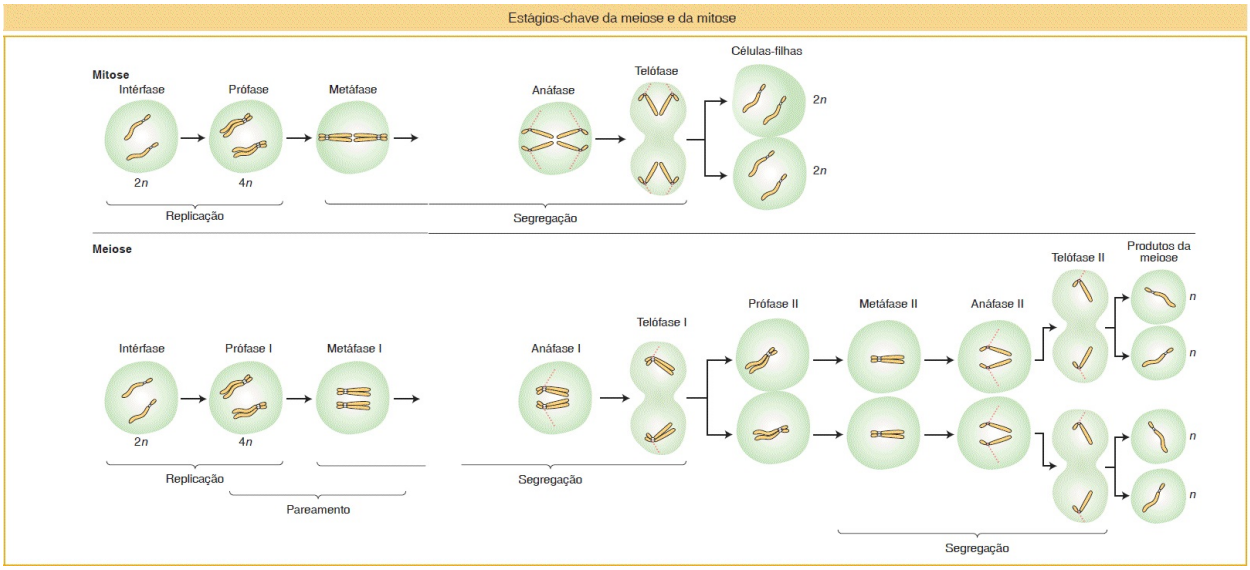


FIGURA 2.8 Representação simplificada da mitose e da meiose em células diploides (2*n* = Diploide; *n* = Haploide). (Versões detalhadas estão demonstradas no Apêndice 2.1, adiante.)

Os bivalentes de todos os cromossomos se movimentam até o equador da célula e, quando a célula se divide, uma díade vai para cada nova célula, puxada por fibras do fuso ligadas próximo aos centrômeros. Na segunda divisão celular da meiose, os centrômeros se dividem e cada membro de uma díade (cada membro de um par de cromátides) se movimenta para uma célula-filha. Portanto, embora o processo tenha início com o mesmo conteúdo genômico da mitose, as duas segregações sucessivas resultam em quatro células haploides. Cada uma das quatro células haploides que constituem os quatro **produtos da meiose** contém um membro de uma tétrade; portanto, o grupo de quatro células por vezes também é denominado uma tétrade. A meiose pode ser resumida como segue:

Início: → dois homólogos

Replicação: → duas díades

Pareamento: → tétrade

Primeira divisão: → uma díade para cada célula-filha

Segunda divisão: → uma cromátide para cada célula-filha

Pesquisas em biologia celular demonstraram que as fibras do fuso que separam os cromossomos são polímeros da molécula tubulina. A separação é causada principalmente por uma despolimerização e consequente encurtamento das fibras no ponto em que elas estão unidas aos cromossomos.

O comportamento dos cromossomos durante a meiose explica claramente a lei de Mendel sobre a segregação igual. Considere um heterozigoto de tipo geral *A*/*a*. Podemos simplesmente seguir o resumo precedente enquanto consideramos o que ocorre com os alelos desse gene:

Início: um homólogo carrega *A* e um carrega *a*

Replicação: uma díade é *AA* e uma é *aa*

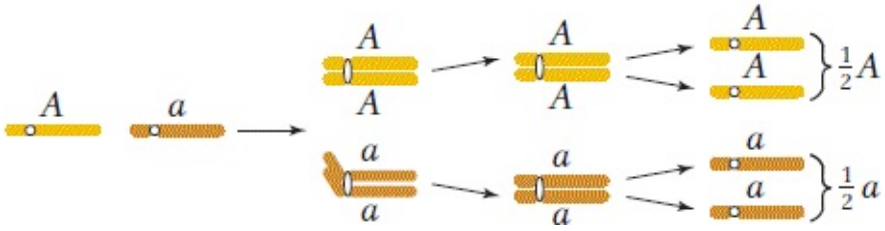
Pareamento: a tétrade é *A*/*A*/*a*/*a*

Produtos da primeira divisão: uma célula *AA*, a outra célula *aa* (o *crossing over* pode misturar esses tipos de produtos, mas a proporção geral não é alterada)

Produtos da segunda divisão: quatro células, duas do tipo *A* e duas do tipo *a*

Portanto, os produtos da meiose de um meiócito heterozigoto *A*/*a* são $\frac{1}{2}$ *A* e $\frac{1}{2}$ *a*, precisamente a proporção igual que é necessária para explicar a primeira lei de Mendel.

Meiose:



Observe que enfocamos os amplos aspectos genéticos da meiose necessários para explicar a herança monogênica. Descrições mais completas dos estágios

detalhados da mitose e da meiose são apresentadas nos Apêndices 2.1 e 2.2 ao fim deste capítulo.

Herança monogênica em haploides

Vimos que a base celular da lei da segregação igual é a segregação dos cromossomos na primeira divisão da meiose. Até então na discussão, a evidência em relação à segregação igual de alelos nos meiócitos de plantas e de animais é *indireta*, com base na observação que cruzamentos demonstram proporções adequadas de progênie esperadas pela segregação igual. Reconheça que os gametas nesses estudos (tais como o de Mendel) precisam advir de *muitos* meiócitos *diferentes*. Entretanto, em alguns organismos, seu ciclo de vida especial possibilita o exame dos produtos de um único meiócito. Esses organismos são denominados haploides, bons exemplos dos quais são a maior parte dos fungos e das algas. Eles passam a maior parte de suas vidas no estado haploide, mas conseguem se cruzar, no processo que forma uma célula diploide temporária, que se torna o meiócito. Em algumas espécies, os quatro produtos de uma única meiose são temporariamente mantidos juntos em um tipo de saco.

Saccharomyces cerevisiae (um fungo), fornece um bom exemplo (ver Organismo-modelo, Levedura, no Capítulo 12). Em fungos, existem formas simples de sexos, denominadas *tipos reprodutivos*. Em *S. cerevisiae*, existem dois tipos reprodutivos e um cruzamento de sucesso somente pode ocorrer entre linhagens de tipos reprodutivos opostos. Vejamos um cruzamento que inclui um mutante de levedura. Colônias de levedura do tipo selvagem normais são brancas, mas, ocasionalmente, surgem mutantes vermelhos em virtude de uma mutação em um gene na via bioquímica que sintetiza a adenina. Utilizemos o mutante vermelho para investigar a segregação igual em um único meiócito. Podemos denominar o alelo mutante *r* para *vermelho*. Qual símbolo podemos utilizar para o alelo normal, ou do tipo selvagem? Em genética experimental, o alelo do tipo selvagem em relação a qualquer gene em geral é designado por um sinal de mais, +. Esse sinal é unido como um sobrescrito ao símbolo inventado para o alelo mutante. Portanto, o alelo do tipo selvagem nesse exemplo será designado r^+, mas com frequência é utilizado apenas o sinal + como uma abreviação. Para verificar a

segregação monogênica, o mutante vermelho é cruzado com o tipo selvagem. O cruzamento será:

$$r^+ \times r$$

Quando duas células do tipo reprodutivo oposto se fundem, é formada uma célula diploide, e é essa célula que se torna o meiócito. No presente exemplo, o meiócito diploide será heterozigoto, r^+/r. A replicação e a segregação de r^+ e r originam uma tétrade de dois produtos meióticos (esporos) de genótipo r^+ e dois de genótipo r, todos contidos dentro de um saco membranoso denominado **asco**. Portanto,

$$r^+/r \rightarrow \left.\begin{matrix} r^+ \\ r^+ \\ r \\ r \end{matrix}\right\} \text{Tétrade no asco}$$

Os detalhes do processo estão demonstrados na Figura 2.9. Se os quatro esporos de um asco forem isolados (representando uma tétrade de cromátides) e utilizados para gerar quatro culturas de leveduras, a segregação igual dentro de um meiócito é revelada diretamente como duas culturas brancas e duas vermelhas. Se analisarmos os esporos aleatórios de muitos meiócitos, encontraremos aproximadamente 50% de vermelhos e 50% de brancos.

Observe a simplicidade da genética haploide: um cruzamento requer a análise de apenas uma meiose; contrariamente, um cruzamento diploide requer uma consideração da meiose em ambos os genitores dos sexos masculino e feminino. Essa simplicidade é um motivo importante para a utilização de haploides como organismos-modelo. Outro motivo é que, em haploides, todos os alelos estão expressos no fenótipo, tendo em vista que não existe mascaramento de recessivos por alelos dominantes no outro homólogo.

Demonstração da segregação igual dentro de um meiócito na levedura *S. cerevisiae*

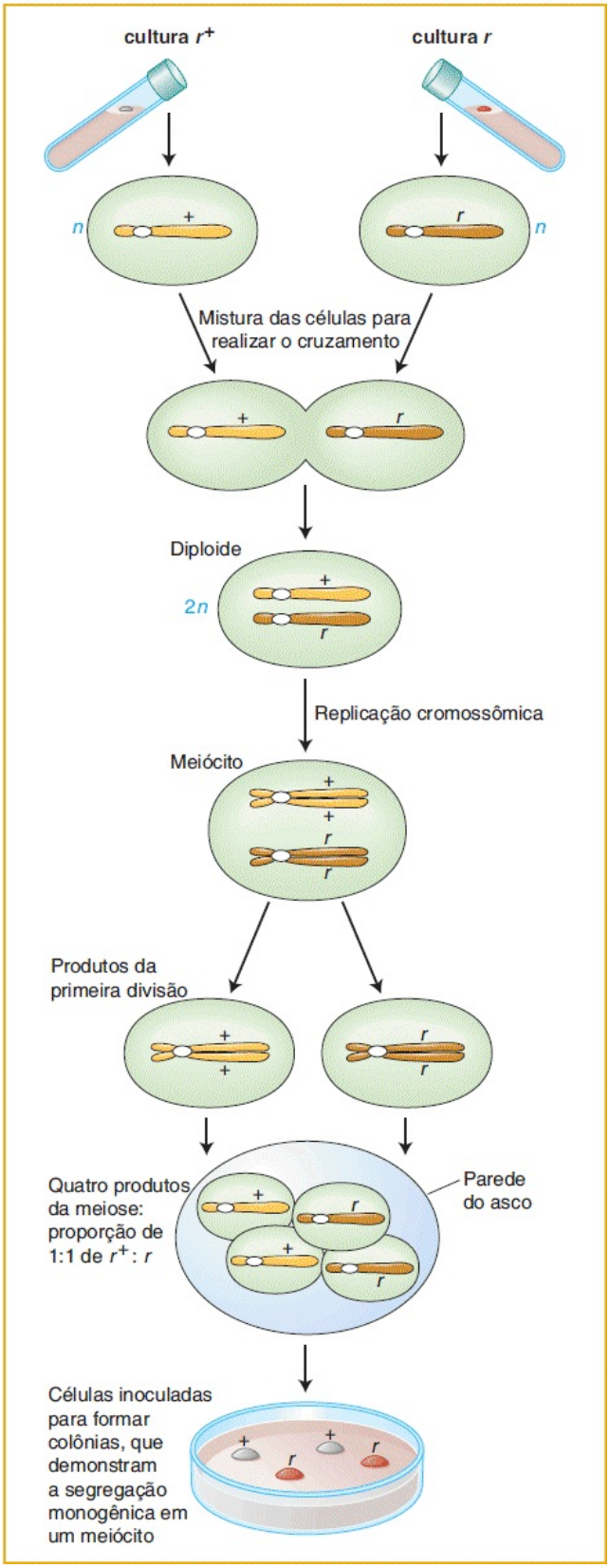


FIGURA 2.9 Um asco isolado do cruzamento + × *r* leva a duas culturas de + e duas culturas de *r*.

2.3 Base molecular dos padrões de herança mendeliana

É claro que Mendel não fazia ideia da natureza molecular dos conceitos com os quais estava trabalhando. Nesta seção, podemos iniciar a introdução de alguns dos conceitos de Mendel em um contexto molecular. Iniciemos com os alelos. Utilizamos o conceito de *alelos* sem defini-los no nível molecular. Quais são as *diferenças estruturais* entre os alelos do tipo selvagem e mutantes no nível do DNA de um gene? Quais são as *diferenças funcionais* no nível proteico? Os alelos mutantes podem ser utilizados para estudar a herança monogênica sem a necessidade de compreender a sua natureza estrutural ou funcional. Entretanto, tendo em vista que um motivo primário para o embarque na herança monogênica é finalmente investigar a função de um gene, devemos compreender a natureza molecular dos alelos do tipo selvagem e mutante em ambos os níveis, estrutural e funcional.

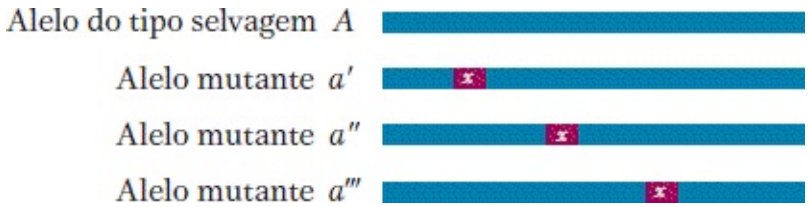
Diferenças estruturais entre os alelos no nível molecular

Mendel propôs que os genes se apresentam em diferentes formas, que atualmente denominamos alelos. O que são alelos no nível molecular? Quando alelos tais como *A* e *a* são examinados no nível do DNA por meio da utilização da tecnologia moderna, em geral observa-se que são idênticos na maior parte de suas sequências e que diferem apenas em um ou diversos nucleotídios das centenas ou dos milhares de nucleotídios que compõem o gene. Portanto, observamos que os alelos são versões verdadeiramente diferentes do mesmo gene. O diagrama a seguir representa o DNA de dois alelos de um gene; a letra x representa uma diferença na sequência de nucleotídios:

Se a sequência de nucleotídios de um alelo é alterada como resultado de um

“acidente” químico raro, é criado um novo alelo mutante. As referidas alterações podem ocorrer em qualquer local ao longo da sequência de nucleotídios de um gene. Por exemplo, uma mutação pode ser uma alteração na identidade de um nucleotídio único, ou a deleção de um ou mais nucleotídios, ou até mesmo a adição de um ou mais nucleotídios.

Existem muitos modos por meio dos quais um gene pode ser alterado pela mutação. Por um lado, a lesão mutacional pode ocorrer em qualquer um de muitos locais diferentes. Podemos representar a situação como segue, em que o azul-escuro indica a sequência de DNA do tipo selvagem normal e o vermelho com a letra x representa a sequência alterada:



Aspectos moleculares da transmissão dos genes

Replicação dos alelos durante a fase S. O que ocorre com os alelos no nível molecular durante a divisão celular? Sabemos que o componente genômico primário de cada cromossomo é uma molécula de DNA. Essa molécula de DNA é replicada durante a fase S, que precede a mitose e a meiose. Conforme veremos no Capítulo 7, a replicação é um processo preciso e, assim, todas as informações genéticas são duplicadas, sejam do tipo selvagem ou mutante. Por exemplo, se uma mutação for o resultado de uma alteração em um único par de nucleotídios — digamos, de GC (tipo selvagem) para AT (mutante) — então em um heterozigoto, a replicação será como segue:

$$\text{homólogo GC} \rightarrow \text{replicação} \rightarrow \begin{matrix}\text{cromátide GC} \\ \text{cromátide GC}\end{matrix}$$

$$\text{homólogo AT} \rightarrow \text{replicação} \rightarrow \begin{matrix}\text{cromátide AT} \\ \text{cromátide AT}\end{matrix}$$

A replicação do DNA antes da mitose em um haploide e um diploide está demonstrada na Figura 2.10. Esse tipo de ilustração serve para nos lembrar que,

em nossas considerações sobre os mecanismos da herança, essencialmente são as moléculas de DNA que estão sendo movimentadas nas células em divisão.

Meiose e mitose no nível molecular. A replicação do DNA durante a fase S produz duas cópias de cada alelo, *A* e *a*, que agora podem ser segregadas em células separadas. A divisão nuclear visualizada no nível do DNA está demonstrada na Figura 2.11.

Demonstração da segregação cromossômica no nível molecular. Interpretamos os padrões de herança fenotípica monogênica em relação à segregação do DNA cromossômico na meiose. Existe algum modo para demonstrar a segregação do DNA diretamente (contrariamente à segregação fenotípica)? A abordagem mais direta seria sequenciar os alelos (digamos, *A* e *a*) nos genitores e nos produtos meióticos: o resultado seria que metade dos produtos apresentaria a sequência do DNA *A* e metade apresentaria a sequência do DNA *a*. O mesmo seria verdadeiro em relação a qualquer sequência de DNA que diferisse nos cromossomos herdados, incluindo aquelas não necessariamente dentro dos alelos correlacionados com fenótipos conhecidos, tais como flores vermelhas e brancas. Portanto, observamos que as regras da segregação anunciadas por Mendel se aplicam não somente aos genes, mas também a qualquer trecho de DNA ao longo de um cromossomo.

As moléculas de DNA se replicam para formar cromátides idênticas

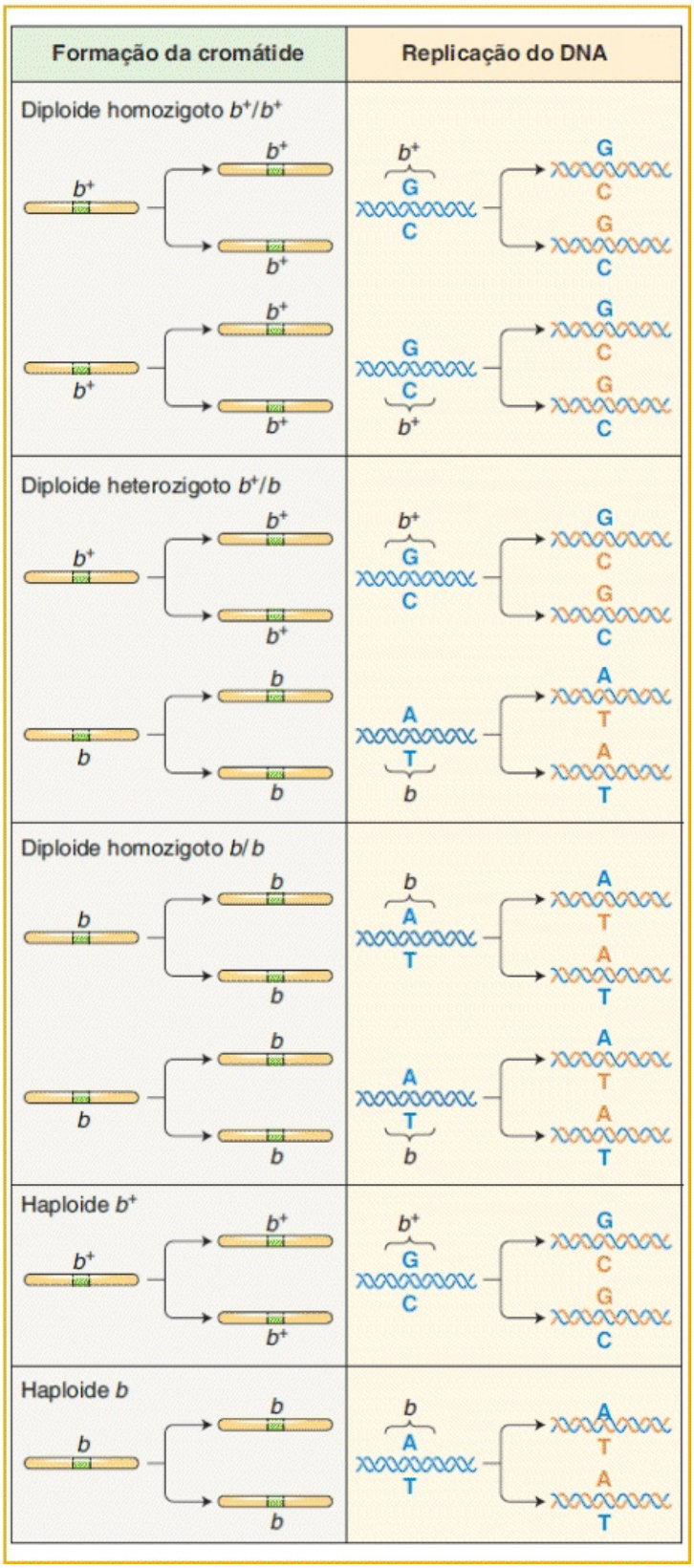


FIGURA 2.10 Cada cromossomo se divide longitudinalmente em duas cromátides (*esquerda*); no nível molecular (*direita*), a molécula de DNA única de cada cromossomo se replica, produzindo duas moléculas de DNA, uma para cada cromátide. Também estão demonstradas diversas combinações de um gene com o alelo do tipo selvagem b^+ e o tipo mutante *b*, causado pela alteração em um único par de bases, de GC para AT. Observe que, no nível do DNA, as duas cromátides produzidas quando um cromossomo se replica são sempre idênticas entre si e ao cromossomo original.

> CONCEITO-CHAVE A herança mendeliana é demonstrada por qualquer segmento de DNA em um cromossomo: por meio dos genes e de seus alelos e por meio de marcadores moleculares não necessariamente associados a qualquer função biológica.

Alelos no nível molecular

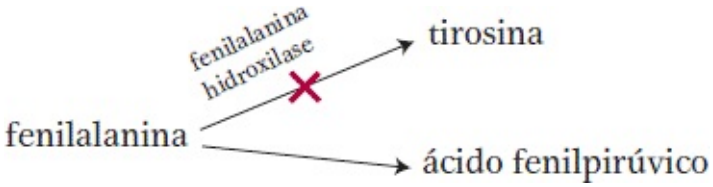
No nível molecular, o fenótipo primário de um gene é a proteína que ele produz. Quais são as diferenças funcionais entre as proteínas que explicam os distintos efeitos dos alelos do tipo selvagem e mutantes sobre as propriedades de um organismo?

Exploremos o tópico ao utilizar a doença humana fenilcetonúria (PKU). Veremos em uma seção posterior na análise de heredogramas que o fenótipo da PKU é herdado como um traço mendeliano recessivo. A doença é causada por um alelo defeituoso do gene que codifica a enzima hepática fenilalanina hidroxilase (PAH). Essa enzima normalmente converte a fenilalanina nos alimentos no aminoácido tirosina:

$$\text{fenilalanina} \xrightarrow[\text{hidroxilase}]{\text{fenilalanina}} \text{tirosina}$$

Entretanto, uma mutação no gene que codifica essa enzima pode alterar a sequência de aminoácidos próxima ao sítio ativo da enzima. Nesse caso, a enzima não consegue se ligar à fenilalanina (seu substrato) ou convertê-la em tirosina. Portanto, a fenilalanina se acumula no corpo e é convertida, em vez disso, em ácido fenilpirúvico. Esse composto interfere no desenvolvimento do sistema

nervoso, levando ao retardo mental.



Os bebês atualmente são testados de modo rotineiro ao nascimento em relação a essa deficiência de processamento. Se a deficiência for detectada, o acúmulo de fenilalanina poderá ser evitado com a utilização de uma dieta especial e o desenvolvimento da doença será interrompido.

A enzima PAH é composta por um único tipo de proteína. Quais alterações ocorreram no DNA mutado do gene da PKU, e como a referida alteração no nível do DNA afeta a função da proteína e produz o fenótipo da doença? O sequenciamento dos alelos mutantes de muitos pacientes com PKU revelou uma pletora de mutações em diferentes sítios ao longo do gene, principalmente nas regiões de codificação da proteína, ou éxons; os resultados estão resumidos na Figura 2.12. Elas representam uma variação de alterações do DNA, mas a maior parte é de pequenas alterações que afetam apenas um par de nucleotídios entre os milhares que constituem o gene. O que esses alelos têm em comum é que eles codificam uma proteína defeituosa, que deixa de apresentar a atividade normal da PAH. Ao alterar um ou mais aminoácidos, todas as mutações inativam alguma parte essencial da proteína codificada pelo gene. O efeito da mutação sobre a função gênica depende da região dentro do gene no qual ocorre a mutação. Uma região funcional importante do gene é aquela que codifica um sítio ativo da enzima; assim, essa região é muito sensível à mutação. Além disso, se observa que uma minoria das mutações ocorre em íntrons, e essas mutações com frequência impedem o processamento normal do transcrito primário do RNA. Algumas das consequências gerais da mutação no nível da proteína estão demonstradas na Figura 2.13. Muitos dos alelos mutantes são de um tipo em geral denominado **alelos nulos**: as proteínas codificadas por eles apresentam ausência completa da função da PAH. Outros alelos mutantes reduzem o nível da função enzimática; por vezes eles são denominados **mutações hipomórficas (*leaky*)**,

tendo em vista que alguma função do tipo selvagem parece “vazar” no fenótipo mutante. O sequenciamento do DNA com frequência detecta alterações que não causam impacto funcional, de modo que elas são funcionalmente do tipo selvagem. Portanto, observamos que os termos *tipo selvagem* e *mutante* por vezes devem ser utilizados com cautela.

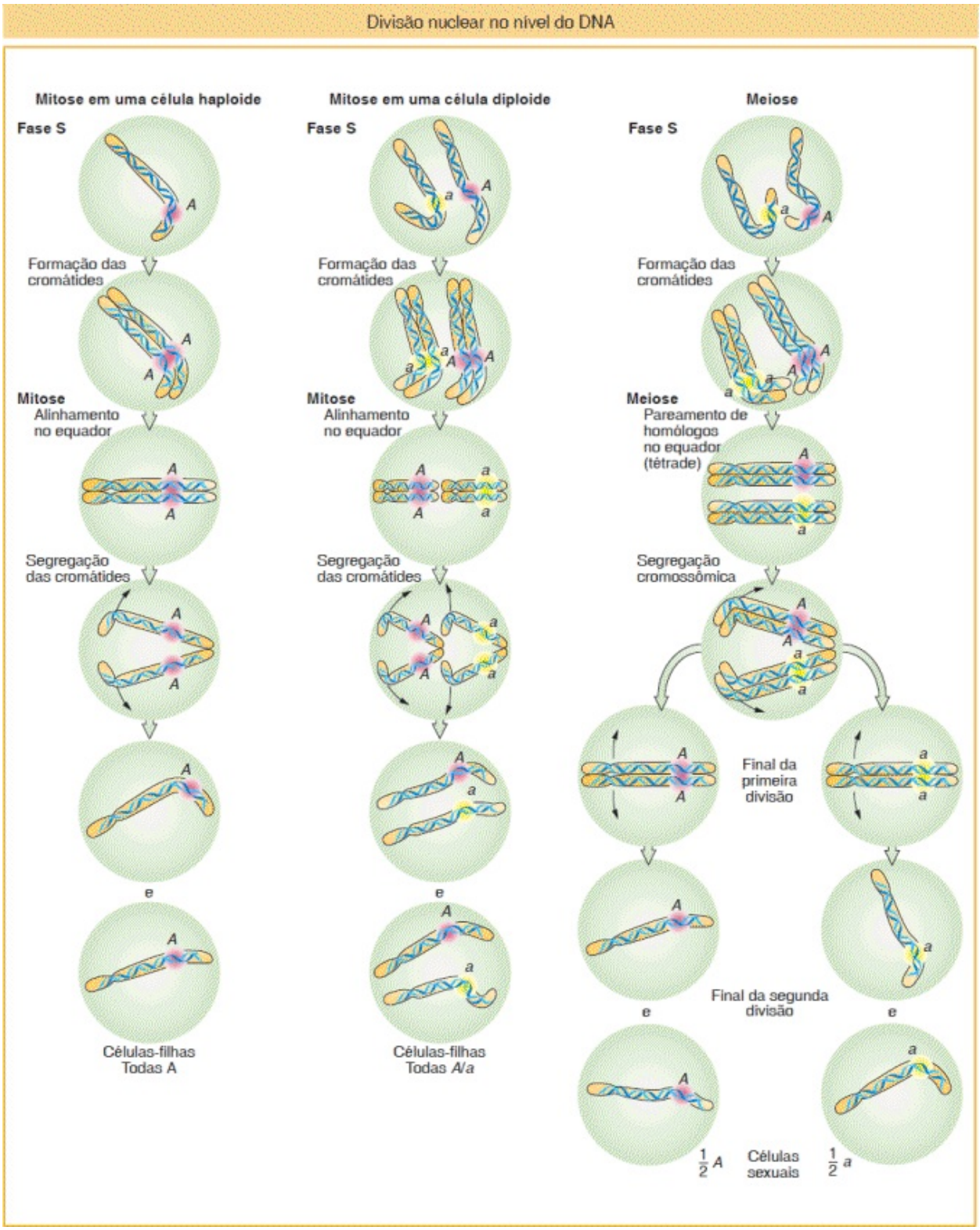


FIGURA 2.11 Transmissão gênica e de DNA na mitose e na meiose em eucariotos. Estão demonstrados a fase S e os principais estágios da mitose e da meiose. As divisões mitóticas (*à esquerda e na parte central*) conservam o genótipo da célula original. À direita, as duas divisões meióticas sucessivas que ocorrem durante

o estágio sexuado do ciclo de vida apresentam o efeito líquido de dividir pela metade o número de cromossomos. Os alelos *A* e *a* de um gene são utilizados para demonstrar como os genótipos são transmitidos na divisão celular.

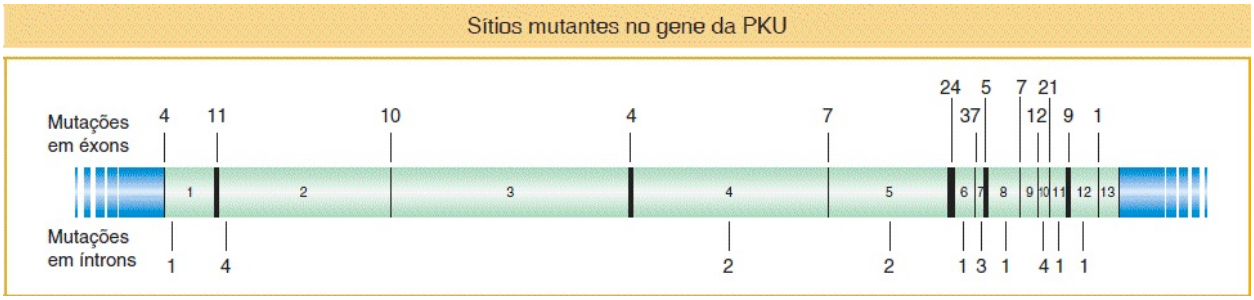


FIGURA 2.12 Muitas mutações no gene da fenilalanina hidroxilase humana que causam mau funcionamento enzimático são conhecidas. O número de mutações nos éxons, ou regiões codificadoras da proteína (*pretas*), está listado acima do gene. O número de mutações nas regiões intrônicas (*verdes, numeradas 1 a 13*) que alteram a recomposição do mRNA está listado abaixo do gene. (Dados de C. R. Scriver, *Ann. Rev. Genet.* 28, 1994, 141-165.)

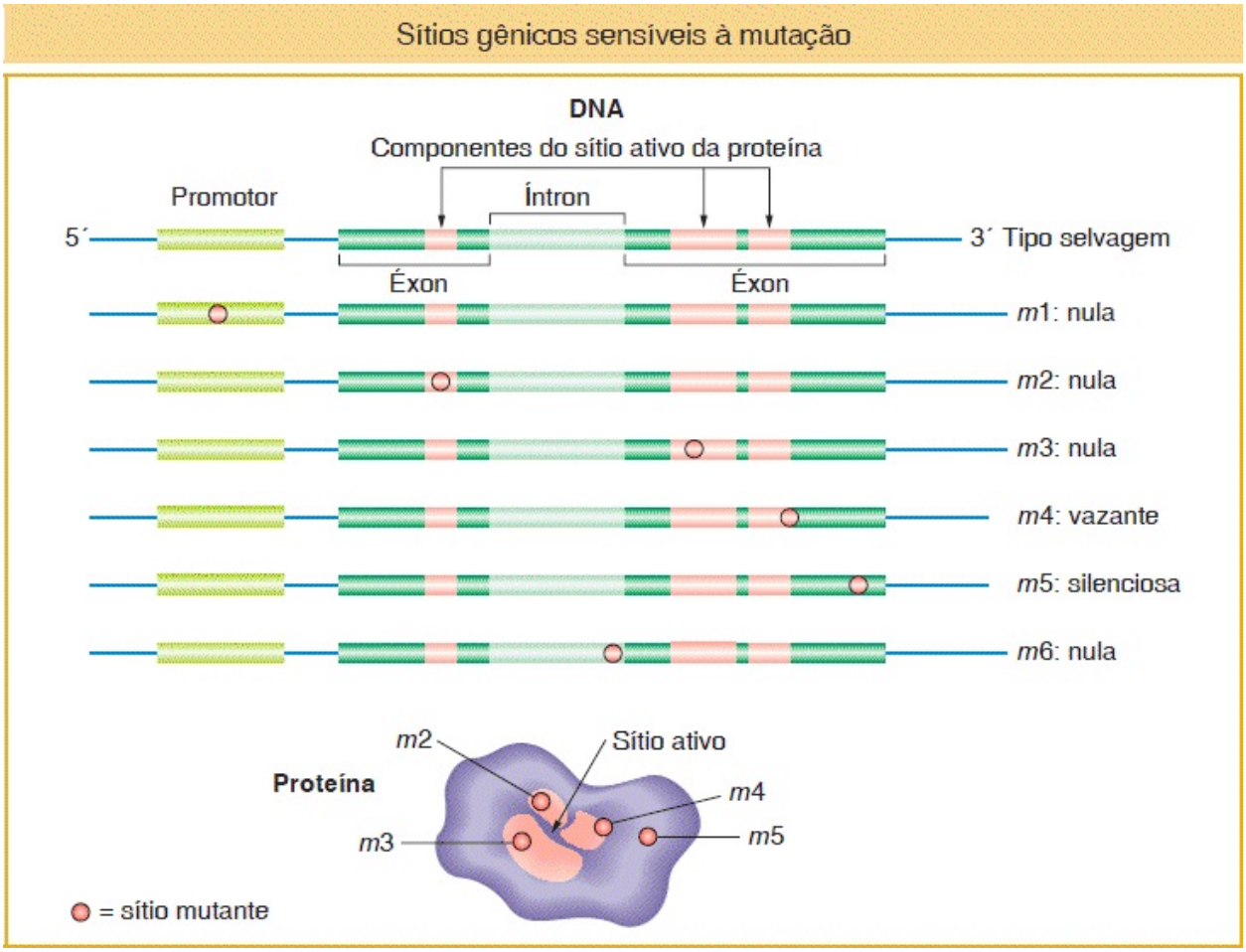


FIGURA 2.13 Mutações nas regiões de um gene que codificam sítios enzimáticos ativos levam à formação de enzimas que não funcionam (mutações nulas). Mutações em outras regiões gênicas podem não apresentar efeitos sobre a função enzimática (mutações silenciosas). Os promotores são sítios importantes na iniciação da transcrição.

> CONCEITO-CHAVE A maior parte das mutações que alteram o fenótipo altera a sequência de aminoácidos do produto proteico do gene, resultando em redução ou ausência de função.

Temos seguido a ideia de que descobrir um conjunto de genes que interferem com a propriedade biológica em investigação é um objetivo importante da genética, tendo em vista que ele define os componentes do sistema. Entretanto, descobrir o modo *preciso* por meio do qual os alelos mutantes levam aos fenótipos mutantes com frequência é desafiador, requerendo não apenas a identificação dos produtos proteicos desses genes, mas também estudos celulares e fisiológicos detalhados para medir os efeitos das mutações. Além disso, descobrir como o conjunto de genes *interage* é um segundo nível de desafio e um tópico que seguiremos posteriormente, com início no Capítulo 6.

Dominância e recessividade. Com uma compreensão a respeito de como os genes atuam por meio de seus produtos proteicos, podemos compreender melhor a dominância e a recessividade. A dominância foi definida anteriormente neste capítulo como o fenótipo mostrado por um heterozigoto. Portanto, formalmente, é o *fenótipo* que é dominante ou recessivo, mas, na prática, os geneticistas com frequência aplicam o termo aos alelos. Essa definição formal não apresenta conteúdo molecular, mas ambas a dominância e a recessividade podem apresentar explicações simples no nível molecular. Introduzimos aqui o tópico, para ser revisitado no Capítulo 6.

Como os alelos podem ser dominantes? Como eles podem ser recessivos? A recessividade é observada em mutações nulas nos genes que são funcionalmente **haplossuficientes**, imprecisamente significando que uma cópia do gene apresenta função suficiente para produzir um fenótipo do tipo selvagem. Embora uma célula

diploide do tipo selvagem normalmente apresente duas cópias totalmente funcionais de um gene, uma cópia de um gene haplossuficiente proporciona produtos genéticos suficientes (em geral uma proteína) para realizar as transações normais da célula. Em um heterozigoto (digamos, +/*m*, em que *m* é um nulo), a única cópia funcional codificada pelo alelo + proporciona produto proteico suficiente para a função celular normal. Em um exemplo simples, presuma que uma célula necessite de no mínimo 10 unidades proteicas para funcionar normalmente. Cada alelo do tipo selvagem consegue produzir 12 unidades. Portanto, um homozigoto do tipo selvagem +/+ produzirá 24 unidades. O heterozigoto +/*m* produzirá 12 unidades, excedendo o mínimo de 10 unidades e, portanto, o alelo mutante é recessivo, tendo em vista que não apresenta impacto sobre o heterozigoto.

Outros genes são **haploinsuficientes**. Nos referidos casos, um alelo mutante nulo será dominante, tendo em vista que, em um heterozigoto (+/*P*), o único alelo do tipo selvagem não consegue proporcionar produto suficiente para a função normal. Como outro exemplo, vamos presumir que a célula necessite de no mínimo 20 unidades dessa proteína, e que o alelo do tipo selvagem produza apenas 12 unidades. Um homozigoto do tipo selvagem +/+ produz 24 unidades, que é superior ao mínimo. Entretanto, um heterozigoto que envolve uma mutação nula (+/*P*) produz apenas 12; portanto, a presença do alelo mutante no heterozigoto resulta em um suprimento inadequado de produtos e se segue um fenótipo mutante.

Em alguns casos, a mutação resulta em uma *nova função* para o gene. As referidas mutações podem ser dominantes, tendo em vista que, em um heterozigoto, o alelo do tipo selvagem não consegue mascarar essa nova função.

A partir das breves considerações anteriores, observamos que o *fenótipo*, a descrição ou a medida que seguimos durante a herança mendeliana, é uma propriedade emergente, com base na natureza dos alelos e no modo como o gene funciona normal ou anormalmente. O mesmo pode ser dito em relação às descrições de dominante e recessivo que aplicamos para um fenótipo.

2.4 Alguns genes descobertos por meio da observação das proporções de segregação

Relembre que um objetivo geral da análise genética atual é dissecar uma propriedade biológica por meio da descoberta do conjunto de genes únicos que a afetam. Aprendemos que um modo importante de identificar esses genes é por meio das proporções de segregação fenotípica geradas por suas mutações — com mais frequência proporções de 1:1 e 3:1, ambas as quais têm por base a segregação igual, conforme definida por Gregor Mendel.

Vejamos alguns exemplos que estendem a abordagem mendeliana para um ambiente experimental moderno. Tipicamente, o pesquisador é confrontado com uma variedade de fenótipos mutantes interessantes, que afetam a propriedade de interesse (tal como aquelas ilustradas na Figura 2.1) e agora precisa saber se eles são herdados como alelos de mutantes únicos. Os alelos mutantes podem ser dominantes ou recessivos, dependendo da sua ação; assim, a questão da dominância também precisa ser considerada na análise.

O procedimento padrão é cruzar o mutante com o tipo selvagem. (Se o mutante for estéril, então é necessária outra abordagem.) Primeiramente, consideraremos três casos simples que abrangem a maior parte dos desfechos possíveis:

1. Uma flor fértil mutante, sem pigmento nas pétalas (p. ex., com pétalas brancas, contrariamente às vermelhas normais).
2. Uma mosca-das-frutas fértil mutante, com asas curtas.
3. Um fungo fértil mutante, que produz excesso de ramos hifais (hiper-ramificação).

Um gene ativo no desenvolvimento da cor da flor

Para iniciar o processo, a planta com flores brancas é cruzada com o tipo selvagem normal vermelho. Todas as plantas da F_1 são com flores vermelhas e, das 500 plantas da F_2 amostradas, 378 são com flores vermelhas e 122 são com flores brancas. Se reconhecermos a existência do erro de amostragem, essas quantidades da F_2 estão muito próximas de uma proporção de $\frac{3}{4}:\frac{1}{4}$, ou 3:1. Tendo

em vista que essa proporção indica a herança monogênica, podemos concluir que o mutante é causado por uma alteração recessiva em um gene único. De acordo com as regras gerais da nomenclatura dos genes, o alelo mutante para as pétalas brancas pode ser denominado *alb*, que se refere a *albino*, e o alelo do tipo selvagem seria alb^+, ou apenas +. (As convenções em relação à nomenclatura dos alelos variam um pouco entre os organismos: algumas das variações estão demonstradas no Apêndice A sobre a nomenclatura.) Conjeturamos que o alelo do tipo selvagem desempenha um papel essencial na produção das pétalas coloridas da planta, uma propriedade que é quase certamente necessária para atrair polinizadores para a flor. O gene poderia estar implicado na síntese bioquímica do pigmento ou na parte do sistema de sinalização que informa às células da flor que iniciem a fabricação do pigmento, ou em uma diversidade de outras possibilidades que necessitam de investigação adicional. No nível puramente genético, os cruzamentos realizados seriam representados simbolicamente como

P	+/+ × *alb*/*alb*
F_1	Todos +/*alb*
F_2	$\frac{1}{4}$ +/+
	$\frac{1}{2}$ +/*alb*
	$\frac{1}{4}$ *alb*/*alb*

ou graficamente conforme nas grades a seguir (ver também Figura 2.5). Esse tipo de grade que demonstra os gametas e as fusões gaméticas é denominado *quadrado de Punnett*, em homenagem a um dos primeiros geneticistas, Reginald C. Punnett. Eles são dispositivos úteis para explicar as proporções genéticas. Encontraremos mais nas discussões posteriores.

P	*alb*	*alb*
+	*+/alb*	*+/alb*
+	*+/alb*	*+/alb*

Todas da F_1 são vermelhas

F_1	+	*alb*
+	*+/+*	*+/alb*
alb	*+/alb*	*alb/alb*

$\frac{3}{4}$ da F_2 são vermelhas,

$\frac{1}{4}$ são brancas

Um gene para o desenvolvimento das asas

No exemplo da mosca-das-frutas, o cruzamento da mosca com asas curtas mutante com o estoque com asas longas do tipo selvagem produziu uma progênie de 788 moscas, classificadas como segue:

196 machos com asas curtas

194 fêmeas com asas curtas

197 machos com asas longas

201 fêmeas com asas longas

No total, existem 390 moscas com asas curtas e 398 com asas longas, muito próximas de uma proporção de 1:1. A proporção é a mesma nos machos e nas fêmeas, novamente dentro dos limites do erro de amostragem. Portanto, a partir desses resultados, o mutante “asas curtas” muito provavelmente foi produzido por

uma mutação dominante. Observe que, para que uma mutação dominante seja expressa, é necessária apenas uma "dose" única do alelo mutante. Assim, na maior parte dos casos, quando o mutante aparecer pela primeira vez na população, ele ocorrerá no estado heterozigoto. (Isso não é verdadeiro em relação a uma mutação recessiva, tal como aquela no exemplo da planta antecedente, que precisa ser homozigota para ser expressa e deve ser originária da autopolinização de uma planta heterozigota não identificada na geração anterior.)

Quando a progênie de asas longas foi intercruzada, todos os descendentes tinham asas longas, conforme esperado de um alelo do tipo selvagem recessivo. Quando a progênie de asas curtas foi intercruzada, seus descendentes demonstraram uma proporção de três quartos curtos para um quarto de asas longas.

As mutações dominantes são representadas por letras ou palavras maiúsculas: no presente exemplo, o alelo mutante pode ser denominado *SH*, que se refere a "curta" (em inglês, *short*). Então, os cruzamentos seriam representados simbolicamente como:

P +/+ × *SH*/+

F_1 $\frac{1}{2}$ +/+

$\frac{1}{2}$ *SH*/+

F_1 +/+ × +/+

Todas +/+

F_1 $\frac{1}{4}$ *SH*/+ × *SH*/+

$\frac{1}{4}$ SH/SH

$\frac{1}{2}$ SH/+

$\frac{1}{4}$ +/+

ou graficamente conforme demonstrado nas grades adiante.

Essa análise do mutante da mosca identifica um gene que faz parte de um subconjunto de genes que, na forma selvagem, são cruciais para o desenvolvimento normal de uma asa. Tal resultado é o ponto de partida para estudos adicionais, cujo foco é identificar o exato desenvolvimento e as vias celulares nas quais o crescimento da asa é interrompido, que, após a identificação, revelariam o momento de ação do alelo selvagem durante o desenvolvimento.

P	+	*SH*
+	+/+	*SH*/+
+	+/+	*SH*/+

F_1	+	+
+	+/+	+/+
+	+/+	+/+

F_1	+	*SH*
+	+/+	*SH*/+
SH	*SH*/+	*SH*/*SH*

Um gene para a ramificação das hifas

Um fungo mutante hiper-ramificado (tal como a colônia semelhante a um botão na Figura 2.1) foi cruzado com um fungo do tipo selvagem com ramificação esparsa normal. Em uma progênie de 300, 152 eram do tipo selvagem e 148 eram hiper-ramificados, muito próximas de uma proporção de 1:1. Inferimos a partir dessa proporção de herança monogênica que a mutação hiper-ramificada é de um gene único. Em haploides, atribuir a dominância normalmente não é possível, mas, para conveniência, podemos denominar o alelo hiper-ramificado *hb* e o tipo selvagem hb^+ ou +. Os cruzamentos obrigatoriamente foram:

P	+ × *hb*
Meiócito diploide	+/*hb*
F_1	$\frac{1}{2}$ +
	$\frac{1}{2}$ *hb*

A análise da mutação e da herança revelou um gene cujo alelo do tipo selvagem é essencial para o controle normal da ramificação, uma função-chave na dispersão do fungo e na aquisição de nutrientes. Agora o mutante precisa ser investigado para saber em que ponto na sequência normal do desenvolvimento ele produz um bloqueio. Essa informação revelará o momento e o local nas células em que o alelo normal atua.

Por vezes, a gravidade de um fenótipo mutante torna o organismo estéril, incapaz de passar pelo ciclo sexual. Como a herança monogênica dos mutantes estéreis pode ser demonstrada? Em um organismo diploide, um mutante recessivo estéril pode ser propagado como um heterozigoto e em seguida o heterozigoto pode ser autofecundado para produzir os esperados 25% de mutantes homozigotos recessivos para o estudo. Um mutante dominante estéril é um beco sem saída genético e não pode ser propagado sexuadamente, mas, em plantas e fungos, tal mutante pode ser facilmente propagado assexuadamente.

E se um cruzamento entre um mutante e um tipo selvagem não produzir uma

proporção de 3:1 ou 1:1, conforme discutido aqui, mas alguma outra proporção? Tal resultado pode ocorrer em virtude das interações de diversos genes ou de um efeito ambiental. Algumas dessas possibilidades são discutidas no Capítulo 6.

Previsão das proporções na progênie ou dos genótipos parentais por meio da aplicação dos princípios da herança monogênica

Podemos resumir a análise da descoberta dos genes como segue:

Observe as proporções fenotípicas na progênie →
Deduza os genótipos dos genitores (*A*/*A*, *A*/*a* ou *a*/*a*)

Entretanto, o mesmo princípio da herança (essencialmente a lei de Mendel sobre a segregação igual) também pode ser utilizado para prever as proporções fenotípicas na progênie de genitores de *genótipos conhecidos*. Esses genitores seriam originários de estoques mantidos pelo pesquisador. Os tipos e as proporções da progênie de cruzamentos tais como *A*/*A* × *A*/*a*, *A*/*A* × *a*/*a*, *A*/*a* × *A*/*a* e *A*/*a* × *a*/*a* podem ser facilmente previstos. Em resumo:

Cruze genitores de genótipos conhecidos →
Preveja as proporções fenotípicas na progênie

Esse tipo de análise é utilizado em cruzamentos gerais para sintetizar genótipos para pesquisas ou para a agricultura. Também é útil para prever probabilidades de diversos desfechos em reproduções humanas em famílias com histórias de doenças monogênicas.

Após a herança monogênica ter sido estabelecida, um indivíduo que apresente o fenótipo dominante, mas de *genótipo desconhecido*, pode ser testado para verificar se o genótipo é homozigoto ou heterozigoto. Um referido teste pode ser realizado por meio do cruzamento do indivíduo (de fenótipo *A*/?) com uma linhagem-testadora recessiva *a*/*a*. Se o indivíduo for heterozigoto, resultará uma proporção de 1:1 ($\frac{1}{2}$ *A*/*a* e $\frac{1}{2}$ *a*/*a*); se o indivíduo for homozigoto, toda a progênie apresentará o fenótipo dominante (todas *A*/*a*). Em geral, o cruzamento de um indivíduo de heterozigosidade desconhecida (em relação a um gene ou mais) com

um genitor totalmente recessivo é denominado um **cruzamento-teste**, e o indivíduo recessivo é denominado **testador**. Encontraremos cruzamentos-teste muitas vezes em todos os capítulos subsequentes; eles são muito úteis para deduzir os eventos meióticos que ocorrem em genótipos mais complexos, tais como di-híbridos e tri-híbridos. A utilização de um testador totalmente recessivo significa que a meiose no genitor testador pode ser ignorada, tendo em vista que todos os seus gametas são recessivos e não contribuem para os fenótipos da progênie. Um teste alternativo em relação à heterozigosidade (útil se o testador recessivo não estiver disponível e o organismo puder ser autofecundado) é simplesmente autofecundar o desconhecido: se o organismo que está sendo testado for heterozigoto, será observada uma proporção de 3:1 na progênie. Os referidos testes são úteis na análise genética de rotina.

> CONCEITO-CHAVE Os princípios da herança (tais como a lei de segregação igual) podem ser aplicados em duas direções: (1) inferindo os genótipos a partir das proporções fenotípicas e (2) prevendo as proporções fenotípicas a partir de genitores de genótipos conhecidos.

2.5 Padrões de herança monogênica ligada ao sexo

Os cromossomos que estivemos analisando até o momento são autossomos, os cromossomos “regulares” que formam a maior parte do conjunto genômico. Entretanto, muitos animais e plantas apresentam um par especial de cromossomos associados ao sexo. Os cromossomos sexuais também se segregam igualmente, mas as proporções fenotípicas observadas na progênie com frequência são diferentes das proporções autossômicas.

Cromossomos sexuais

A maior parte dos animais e muitas plantas demonstram dimorfismo sexual; em outras palavras, os indivíduos são do sexo masculino ou feminino. Na maioria desses casos, o sexo é determinado por um par especial de **cromossomos**

sexuais. Observemos os seres humanos como um exemplo. As células do corpo humano apresentam 46 cromossomos: 22 pares homólogos de autossomos, mais 2 cromossomos sexuais. As mulheres apresentam um par de cromossomos sexuais idênticos denominados **cromossomos X**. Os homens apresentam um par não idêntico, composto por um X e um Y. O **cromossomo Y** é consideravelmente mais curto que o X. Portanto, se deixarmos A representar os cromossomos autossômicos, podemos escrever:

$$\text{Mulheres} = 44A + XX$$
$$\text{Homens} = 44A + XY$$

Na meiose em indivíduos do sexo feminino, os dois cromossomos X pareiam e segregam como os autossomos e, assim, cada ovócito recebe um cromossomo X. Portanto, em relação aos cromossomos sexuais, os gametas são de apenas um tipo e diz-se que o indivíduo do sexo feminino é o **sexo homogamético**. Na meiose em indivíduos do sexo masculino, os cromossomos X e Y pareiam ao longo de uma região curta, que assegura que X e Y se separem, de modo que existem dois tipos de espermatozoides, metade com um X e a outra metade com um Y. Portanto, o indivíduo do sexo masculino é denominado **sexo heterogamético**.

Os padrões de herança de genes nos cromossomos sexuais são diferentes daqueles dos genes autossômicos. Os padrões de herança dos cromossomos sexuais foram investigados pela primeira vez no início da década de 1900 no laboratório do grande geneticista Thomas Hunt Morgan, com a utilização da mosca-das-frutas *Drosophila melanogaster* (ver Organismo-modelo, adiante). Esse inseto tem sido um dos organismos de pesquisa mais importantes em genética; seu ciclo de vida curto e simples contribui para a sua utilidade nesse sentido. As moscas-das-frutas apresentam três pares de autossomos, mais um par de cromossomos sexuais, novamente denominados X e Y. Assim como em mamíferos, as fêmeas da *Drosophila* apresentam a constituição XX e os machos são XY. Entretanto, o mecanismo de determinação sexual na *Drosophila* difere daquele em mamíferos. Na *Drosophila*, o *número de cromossomos X* em relação aos autossomos determina o sexo: dois X resultam em uma fêmea e um X resulta

em um macho. Em mamíferos, a *presença do cromossomo Y* determina a masculinidade e a ausência de um Y determina a feminilidade. Entretanto, é importante observar que, apesar dessa base um pouco diferente para a determinação sexual, os padrões de herança monogênica dos genes nos cromossomos sexuais são extraordinariamente semelhantes na *Drosophila* e em mamíferos.

As plantas vasculares demonstram uma diversidade de arranjos sexuais. As **espécies dioicas** são aquelas que demonstram dimorfismo sexual semelhante ao dos animais, com as plantas do sexo feminino apresentando flores que contêm apenas ovários e as plantas do sexo masculino apresentando flores que contêm apenas anteras (Figura 2.14). Algumas das plantas dioicas, mas não todas, apresentam um par não idêntico de cromossomos associado ao (e quase certamente determinando o) sexo da planta. Das espécies com cromossomos sexuais não idênticos, uma grande proporção apresenta um sistema XY. Por exemplo, a planta dioica *Melandrium album* apresenta 22 cromossomos por célula: 20 autossomos, mais 2 cromossomos sexuais, com fêmeas XX e machos XY. Outras plantas dioicas não apresentam pares de cromossomos visivelmente diferentes; elas ainda podem apresentar cromossomos sexuais, mas não de tipos visivelmente distinguíveis.

Padrões de herança ligada ao sexo

A citogenética divide os cromossomos X e Y em regiões homólogas e diferenciais. Novamente, utilizemos os seres humanos como um exemplo (Figura 2.15). As regiões diferenciais, que contêm a maioria dos genes, não apresentam correspondentes no outro cromossomo sexual. Portanto, em homens, os genes nas regiões diferenciais são denominados **hemizigotos** (“meio-zigotos”). A região diferencial do cromossomo X contém muitas centenas de genes; a maior parte desses genes não participa na função sexual, e eles influenciam uma grande variedade de propriedades humanas. O cromossomo Y contém apenas algumas dúzias de genes. Alguns desses genes apresentam correspondentes no cromossomo X, mas a maior parte não. Esse último tipo participa da função sexual masculina. Um desses genes, *SRY*, determina a própria masculinidade.

Diversos outros genes são específicos para a produção de espermatozoides em homens.

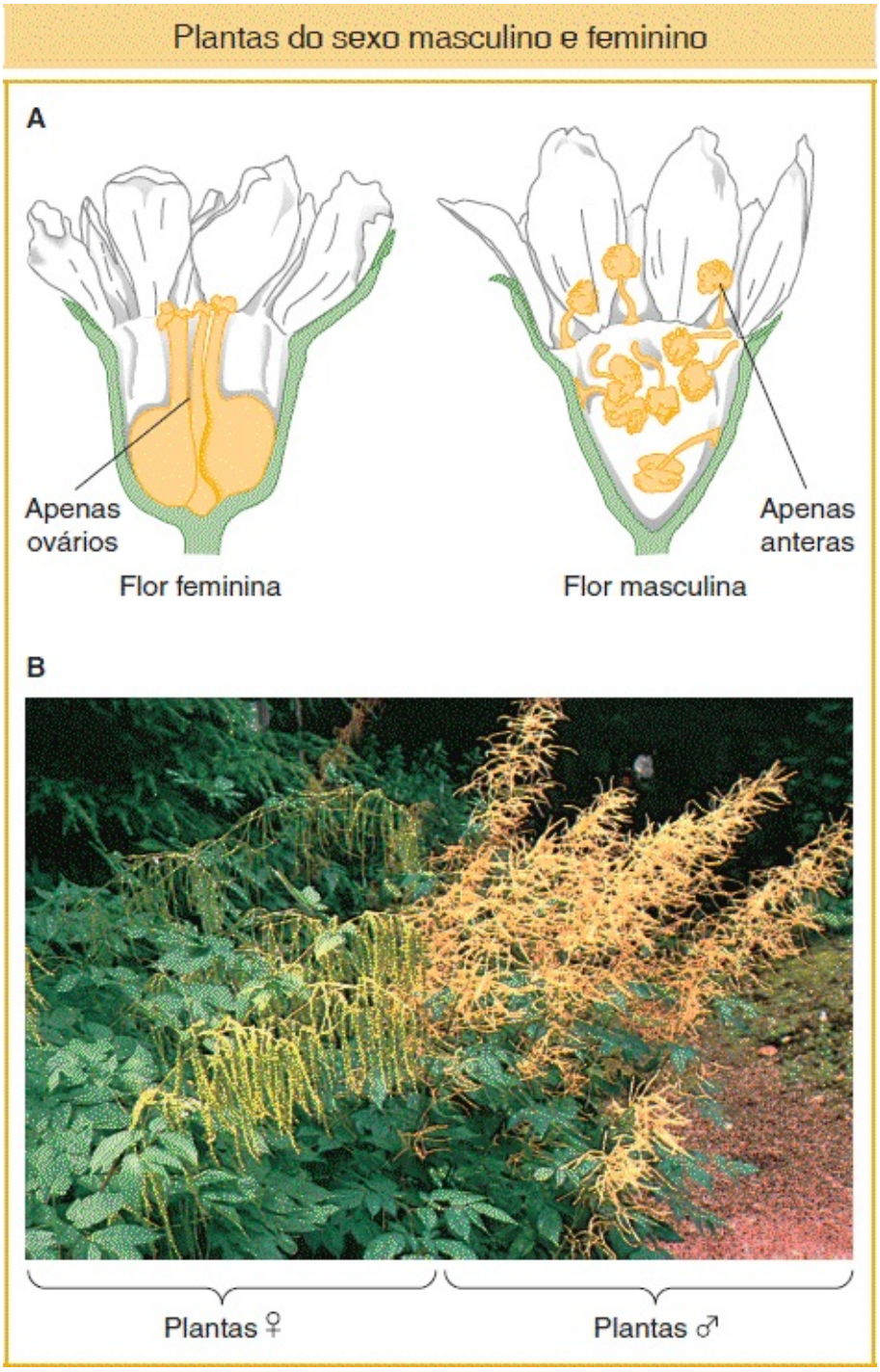


FIGURA 2.14 Exemplos de duas espécies de plantas dioicas são: **A.** *Osmaronia dioica*; **B.** *Aruncus dioicus*. (*A.* Leslie Bohm; *B.* Anthony Griffiths.)

Em geral, diz-se que os genes nas regiões diferenciais demonstram padrões de herança denominados **ligados ao sexo**. Os alelos mutantes na região diferencial do cromossomo X demonstram um padrão de herança monogênica denominado **ligado ao X**. Os alelos mutantes dos poucos genes na região diferencial do cromossomo Y demonstram **ligação ao Y**. Um gene que é ligado ao sexo pode apresentar proporções fenotípicas que são diferentes em cada sexo. Nesse sentido, os padrões de herança ligada ao sexo contrastam com os padrões de herança dos genes nos autossomos, que são os mesmos em cada sexo. Se a localização genômica de um gene for desconhecida, um padrão de herança ligado ao sexo indica que o gene se encontra em um cromossomo sexual.

Os cromossomos X e Y humanos apresentam duas regiões homólogas curtas, uma em cada extremidade (ver Figura 2.15). Como essas regiões são homólogas, elas são semelhantes a regiões autossômicas e, assim, são denominadas **regiões pseudoautossômicas 1** e **2**. Uma dessas regiões ou ambas pareiam na meiose e sofrem *crossing over* (ver Capítulo 4 para detalhes sobre o *crossing over*). Por esse motivo, os cromossomos X e Y podem atuar como um par e segregar em números iguais de espermatozoides.

Herança ligada ao X

Para o nosso primeiro exemplo de ligação ao X, nos voltamos para a cor dos olhos na *Drosophila*. A cor dos olhos do tipo selvagem da *Drosophila* é vermelho-escura, mas existem linhagens puras com olhos brancos (Figura 2.16). Essa diferença fenotípica é determinada por dois alelos de um gene localizado na região diferencial do cromossomo X. O alelo mutante no presente caso é *w* para olhos brancos (a letra minúscula indica que o alelo é recessivo) e o alelo do tipo selvagem correspondente é w^+. Quando machos de olhos brancos são cruzados com fêmeas de olhos vermelhos, toda a progênie da F_1 apresenta olhos vermelhos, sugerindo que o alelo que determina olhos brancos seja recessivo. O cruzamento desses machos e fêmeas da F_1 de olhos vermelhos produz uma proporção da F_2 de 3:1 de moscas de olhos vermelhos para moscas de olhos brancos, mas *todas as moscas de olhos brancos são machos*. Esse padrão de herança, que demonstra uma clara diferença entre os sexos, está explicado na Figura 2.17. A base do

padrão de herança é que todas as moscas da F_1 recebem um alelo do tipo selvagem de suas mães, mas as fêmeas da F_1 também recebem um alelo de olhos brancos de seus pais. Portanto, todas as fêmeas da F_1 são heterozigotas do tipo selvagem (w^+/w), e os machos da F_1 são hemizigotos do tipo selvagem (w^+). As fêmeas da F_1 transmitem o alelo de olhos brancos para metade de seus filhos, que o expressam, e para metade de suas filhas, que não o expressam, certamente porque precisam herdar o alelo do tipo selvagem de seus pais.

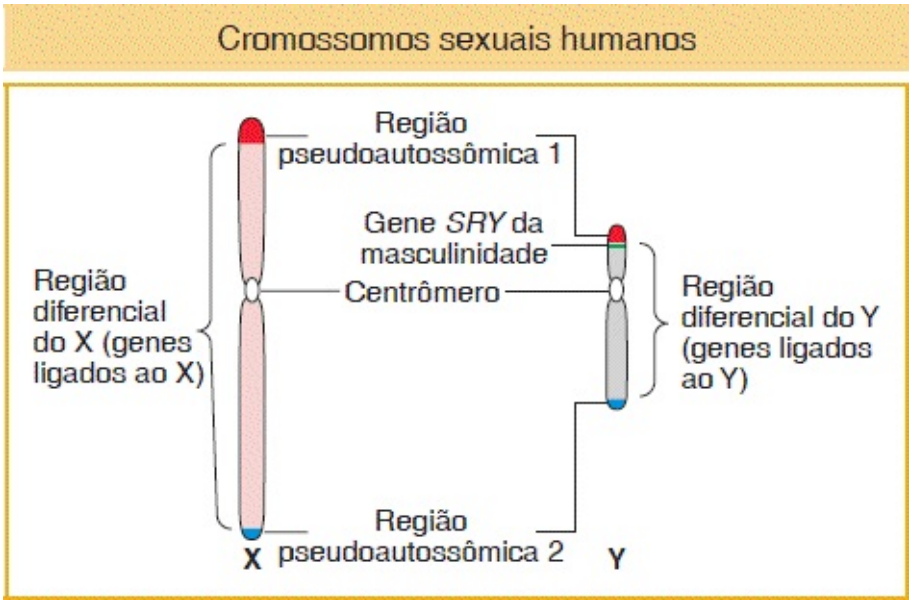


FIGURA 2.15 Os cromossomos sexuais humanos contêm uma região diferencial e duas regiões de pareamento. As regiões foram localizadas ao observar onde os cromossomos pareavam na meiose e onde não pareavam.

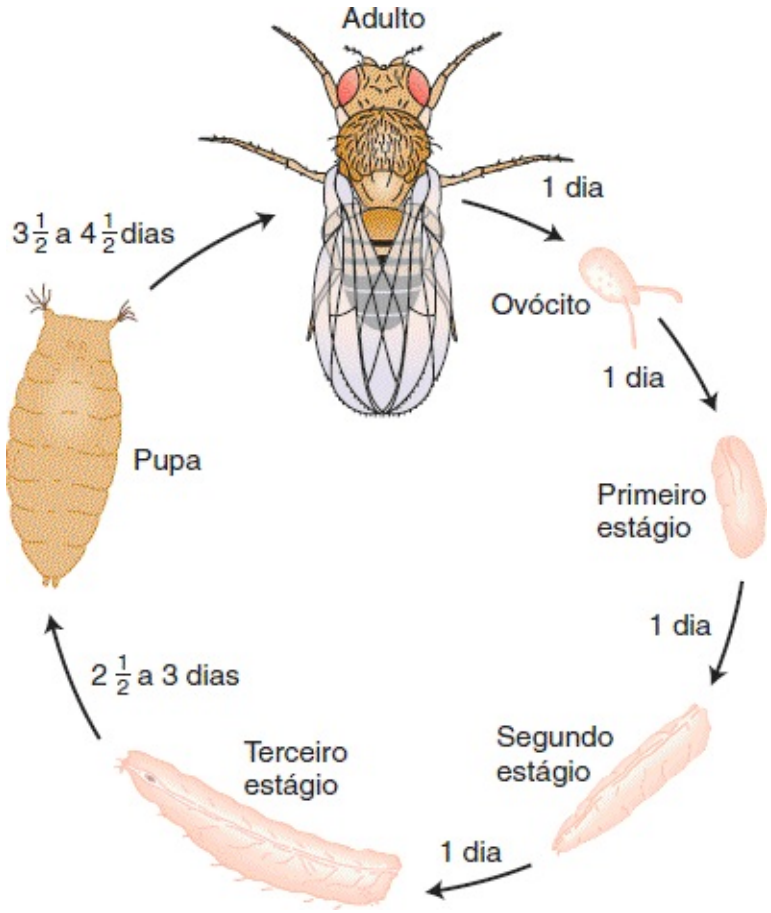
ORGANISMO-MODELO *Drosophila*

A *Drosophila melanogaster* foi um dos primeiros organismos-modelo a serem utilizados em genética. Encontra-se prontamente disponível a partir de frutas maduras, apresenta um ciclo de vida curto, e é de simples cultivo e cruzamento. O sexo é determinado pelos cromossomos sexuais X e Y (XX = Fêmea, XY = Macho), e machos e fêmeas são facilmente distinguidos. Os fenótipos mutantes surgem regularmente em populações de laboratório e sua frequência pode ser aumentada por meio do tratamento com radiação

mutagênica ou substâncias químicas. É um organismo diploide, com quatro pares de cromossomos homólogos (2*n* = 8). Nas glândulas salivares e determinados outros tecidos, ciclos múltiplos de replicação do DNA sem divisão cromossômica resultam em "cromossomos gigantes", cada um com um padrão de bandas único, que proporciona aos geneticistas pontos de referência para o estudo de mapeamento e de rearranjos cromossômicos. Existem muitas espécies e raças de *Drosophila*, que tem sido uma matéria-prima importante para o estudo da evolução.

Drosophila melanogaster, a mosca-das-frutas. (©blickwinkel/Alamy.)

> O tempo voa como uma flecha; moscas-das-frutas gostam de banana.[1]
> (Groucho Marx)



[1]*Time flies like an arrow; fruit flies like a banana.*

O cruzamento recíproco fornece um resultado diferente; ou seja, o cruzamento entre fêmeas com olhos brancos e machos com olhos vermelhos fornece uma F_1 na qual todas as fêmeas têm olhos vermelhos, mas todos os machos têm olhos brancos. Nesse caso, cada fêmea herdou o alelo w^+ dominante do cromossomo X do pai, enquanto cada macho herdou o alelo *w* recessivo de sua mãe. A F_2 é composta por metade de moscas com olhos vermelhos e metade de moscas com olhos brancos de ambos os sexos. Portanto, na ligação ao sexo, observamos exemplos não apenas de diferentes proporções em diferentes sexos, mas também de diferenças entre cruzamentos recíprocos.

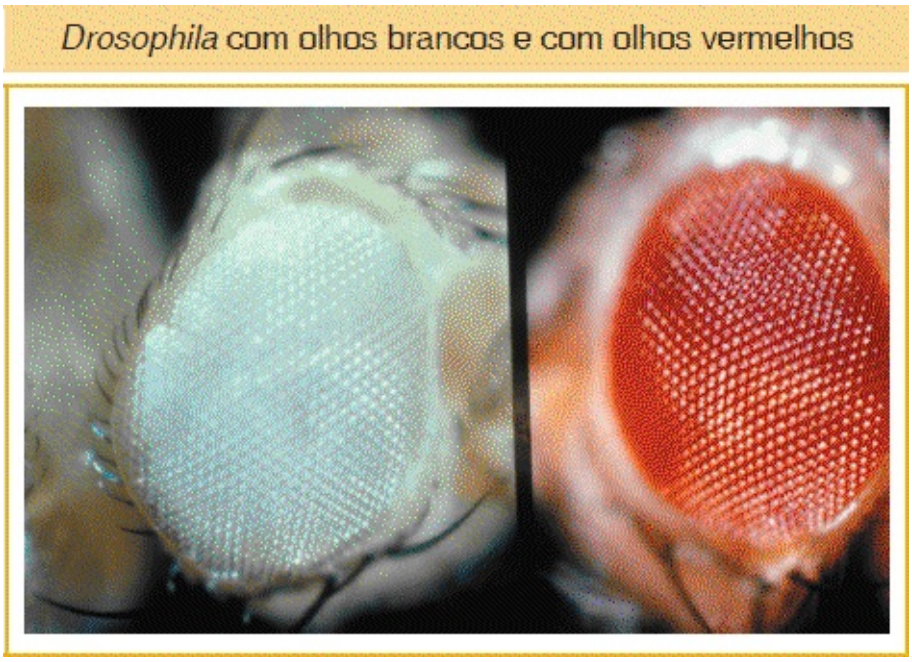


FIGURA 2.16 A mosca com olhos vermelhos é do tipo selvagem e a mosca com olhos brancos é um mutante. (Science Source/Getty Images.)

Observe que a cor dos olhos da *Drosophila* não apresenta correspondência com a determinação do sexo e, assim, temos uma ilustração do princípio de que os genes nos cromossomos sexuais não estão necessariamente relacionados com a função sexual. O mesmo é verdadeiro em seres humanos: na discussão da análise de heredogramas posteriormente neste capítulo, devemos observar muitos genes ligados ao X, ainda que poucos possam ser interpretados como relacionados à função sexual.

O alelo anormal associado à cor dos olhos brancos na *Drosophila* é recessivo, mas também surgem alelos anormais de genes no cromossomo X que são dominantes, tal como o mutante de asas peludas (*Hw*) da *Drosophila*. Nos referidos casos, o alelo do tipo selvagem (Hw^+) é recessivo. Os alelos anormais dominantes demonstram o padrão de herança correspondente àquele do alelo do tipo selvagem para olhos vermelhos no exemplo precedente. As proporções obtidas são as mesmas.

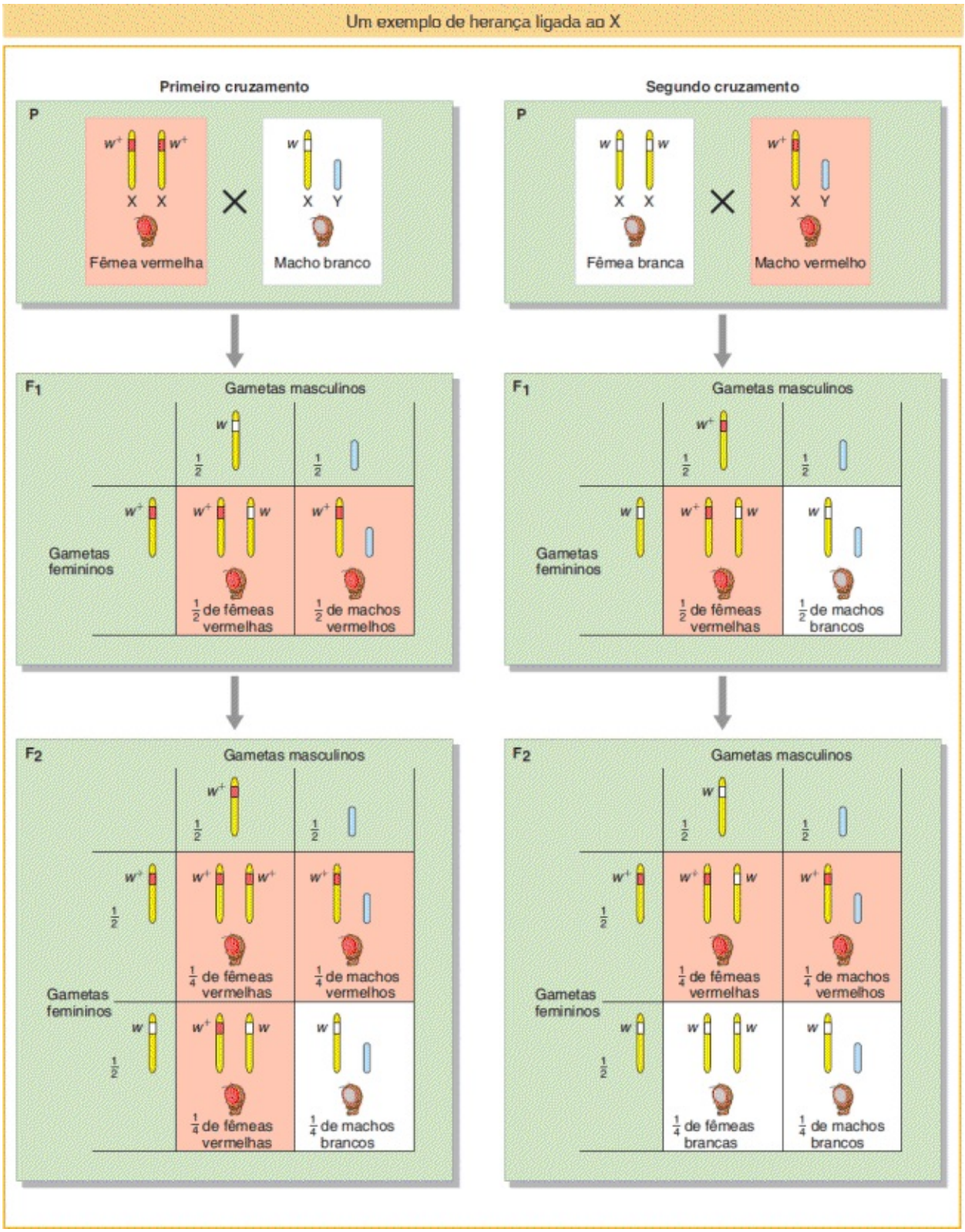


FIGURA 2.17 Cruzamentos recíprocos entre *Drosophila* com olhos vermelhos (*vermelho*) e com olhos brancos (*branco*) fornecem diferentes resultados. Os alelos são ligados ao X, e a herança do cromossomo X

explica as proporções fenotípicas observadas, que são diferentes daquelas dos genes autossômicos. (Na *Drosophila* e em muitos outros sistemas experimentais, é utilizado um sinal de mais sobrescrito para designar o alelo normal, ou do tipo selvagem. Aqui, w^+ codifica os olhos vermelhos e *w* codifica os olhos brancos.)

> CONCEITO-CHAVE A herança ligada ao sexo demonstra regularmente diferentes proporções fenotípicas nos dois sexos da progênie, bem como diferentes proporções em cruzamentos recíprocos.

Historicamente, nas décadas iniciais do século 20, a demonstração por Morgan da herança ligada ao X dos *olhos brancos* na *Drosophila* foi uma peça-chave da evidência que sugeriu que os genes estão, de fato, localizados em cromossomos, tendo em vista que um padrão de herança foi correlacionado com um par específico de cromossomos. A ideia se tornou conhecida como a "teoria cromossômica da herança". Naquele período na história, havia sido recentemente demonstrado que, em muitos organismos, o sexo é determinado por um cromossomo X e um Y e que, em machos, esses cromossomos segregam igualmente na meiose para regenerar quantidades iguais de machos e fêmeas na próxima geração. Morgan reconheceu que a herança dos alelos do gene da cor dos olhos é exatamente paralela à herança dos cromossomos X na meiose; portanto, provavelmente o gene estava no cromossomo X. A herança dos *olhos brancos* foi estendida às linhagens de *Drosophila* que apresentavam números anormais de cromossomos sexuais. Com a utilização dessa nova situação, ainda foi possível prever os padrões de herança de genes a partir da segregação dos cromossomos anormais. Essas previsões terem comprovado estar corretas foi um teste convincente da teoria cromossômica.

Outras análises genéticas revelaram que, em galinhas e mariposas, a herança ligada ao sexo poderia ser explicada apenas se a fêmea fosse o sexo heterogamético. Nesses organismos, os cromossomos sexuais femininos foram designados ZW e os masculinos foram designados ZZ.

2.6 Análise de heredogramas humanos

Os cruzamentos humanos, assim como aqueles dos organismos experimentais, fornecem muitos exemplos de herança monogênica. Entretanto, cruzamentos experimentais controlados não podem ser realizados em seres humanos e, assim, os geneticistas precisam recorrer ao exame de registros médicos na esperança de que tenham ocorrido cruzamentos informativos (tais como cruzamentos de mono-híbridos) que possam ser utilizados para inferir a herança monogênica. Um referido exame dos registros de cruzamentos é denominado **análise de heredogramas**. Um membro de uma família que primeiramente chama a atenção de um geneticista é denominado **probando** (ou propósito). Normalmente, o fenótipo do probando é de algum modo excepcional. Por exemplo, o probando pode apresentar algum tipo de distúrbio clínico. O investigador em seguida traça o histórico do fenótipo por meio do histórico da família e desenha uma árvore genealógica, ou heredograma, com a utilização dos símbolos-padrão fornecidos na Figura 2.18.

Para verificar a herança monogênica, os padrões no heredograma têm de ser interpretados de acordo com a lei de Mendel sobre a segregação igual, mas os seres humanos normalmente têm poucos filhos e, assim, em virtude desse pequeno tamanho da progênie, as proporções esperadas de 3:1 e 1:1 normalmente não são observadas, exceto se muitos heredogramas semelhantes forem combinados. A abordagem para a análise de heredogramas também depende de um dos fenótipos contrastantes ser um distúrbio raro, ou ambos os fenótipos de um par serem comuns (caso em que se diz que são “morfos” de um polimorfismo). A maior parte dos heredogramas é desenhada por motivos clínicos e, portanto, dizem respeito a doenças que são, quase por definição, raras. Nesse caso, temos dois fenótipos: a presença e a ausência do distúrbio. Quatro padrões de herança monogênica são revelados nos heredogramas. Primeiramente, observemos os distúrbios recessivos causados por alelos recessivos de genes autossômicos únicos.

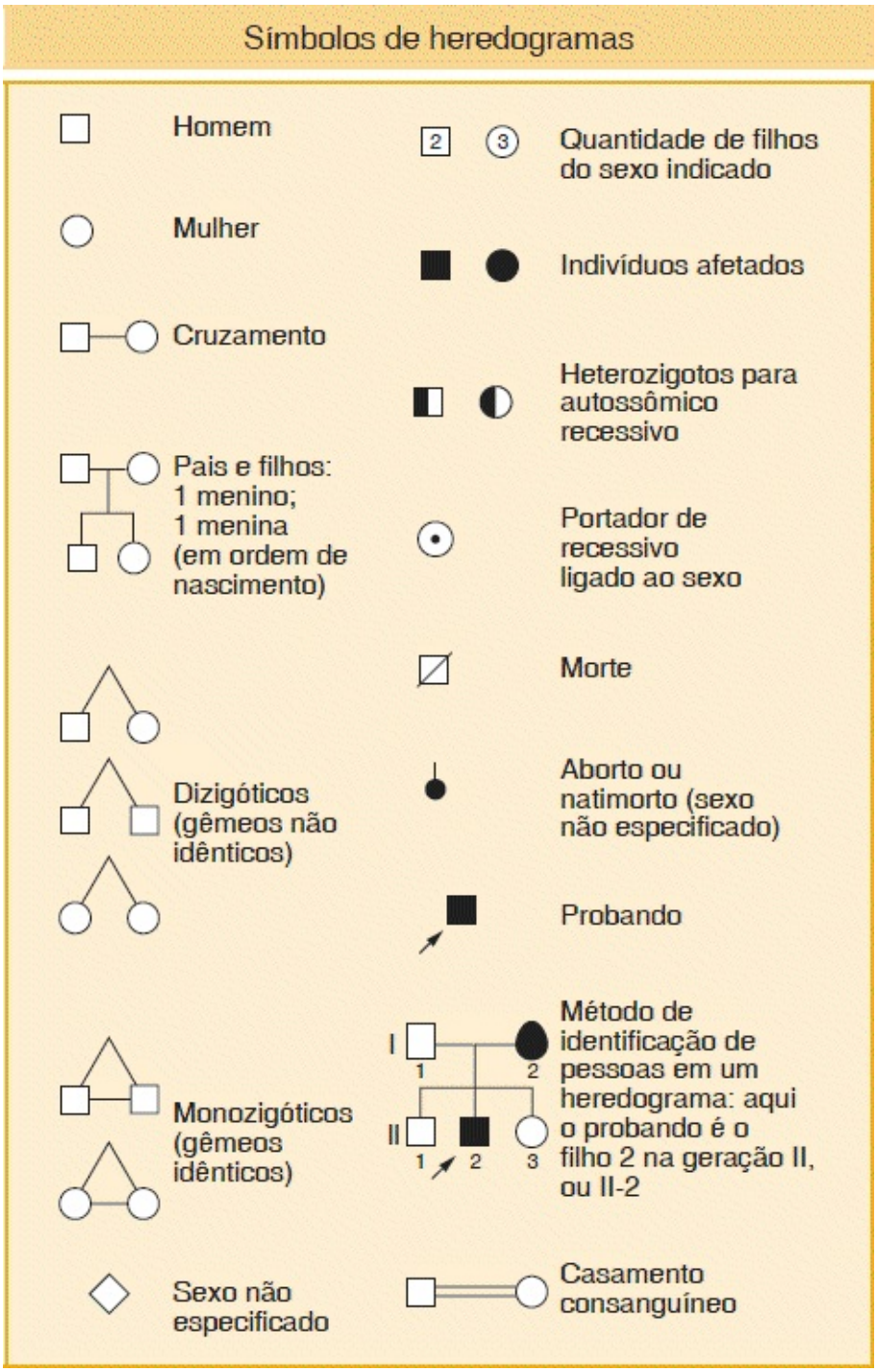


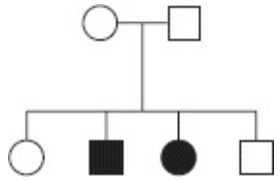
FIGURA 2.18 Uma diversidade de símbolos é utilizada na análise de heredogramas humanos.

Distúrbios autossômicos recessivos

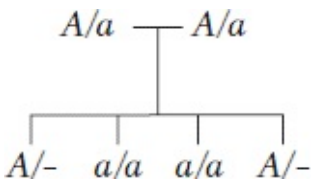
O fenótipo afetado de um distúrbio autossômico recessivo é herdado como um alelo recessivo. Portanto, o fenótipo não afetado correspondente foi

necessariamente herdado como o alelo dominante correspondente. Por exemplo, a doença humana fenilcetonúria (PKU), discutida anteriormente, é herdada de modo mendeliano simples como um fenótipo recessivo, com a PKU determinada pelo alelo *p* e a condição normal determinada por *P*. Portanto, pessoas com essa doença têm genótipo *p*/*p*, e pessoas que não apresentam a doença são *P*/*P* ou *P*/*p*. Relembre que o termo *tipo selvagem* e seus símbolos alélicos não são utilizados na genética humana, tendo em vista que é impossível definir o tipo selvagem.

Quais padrões em um heredograma revelariam a herança autossômica recessiva? Os dois pontos-chave são que (1) em geral o distúrbio aparece na progênie de genitores não afetados e (2) a progênie afetada inclui indivíduos dos sexos masculino e feminino. Quando sabemos que tantos homens quanto mulheres são afetados, podemos inferir que muito provavelmente estamos lidando com a herança mendeliana simples de um gene em um autossomo, em vez de um gene em um cromossomo sexual. O heredograma típico a seguir ilustra o ponto-chave de que crianças afetadas nascem de pais não afetados:



A partir desse padrão, podemos deduzir um cruzamento mono-híbrido simples, com o alelo recessivo responsável pelo fenótipo excepcional (indicado em preto). Ambos os genitores têm de ser heterozigotos — digamos, *A*/*a*; ambos precisam apresentar um alelo *a*, tendo em vista que cada um contribuiu com um alelo *a* para cada criança afetada, e ambos têm de apresentar um alelo *A*, tendo em vista que são fenotipicamente normais. Podemos identificar os genótipos das crianças (demonstrados da esquerda para a direita) como *A*/—, *a*/*a*, *a*/*a* e *A*/—. Portanto, o heredograma pode ser re-escrito como segue:



Esse heredograma não ampara a hipótese de herança recessiva ligada ao X, tendo em vista que, sob essa hipótese, uma filha afetada necessariamente tem a mãe heterozigota (possível) e o pai hemizigoto, o que é claramente impossível, tendo em vista que o pai teria expressado o fenótipo do distúrbio.

Observe que, embora as regras mendelianas estejam atuando, as proporções mendelianas não são necessariamente observadas em famílias isoladas em virtude do pequeno tamanho de amostra, conforme previsto anteriormente. No exemplo precedente, observamos uma proporção fenotípica de 1:1 na progênie de um cruzamento mono-híbrido. Se o casal tivesse, digamos, 20 filhos, a proporção seria algo como 15 filhos não afetados e cinco com PKU (uma proporção de 3:1), mas em uma amostra pequena de quatro filhos, qualquer proporção é possível e todas as proporções são comumente encontradas.

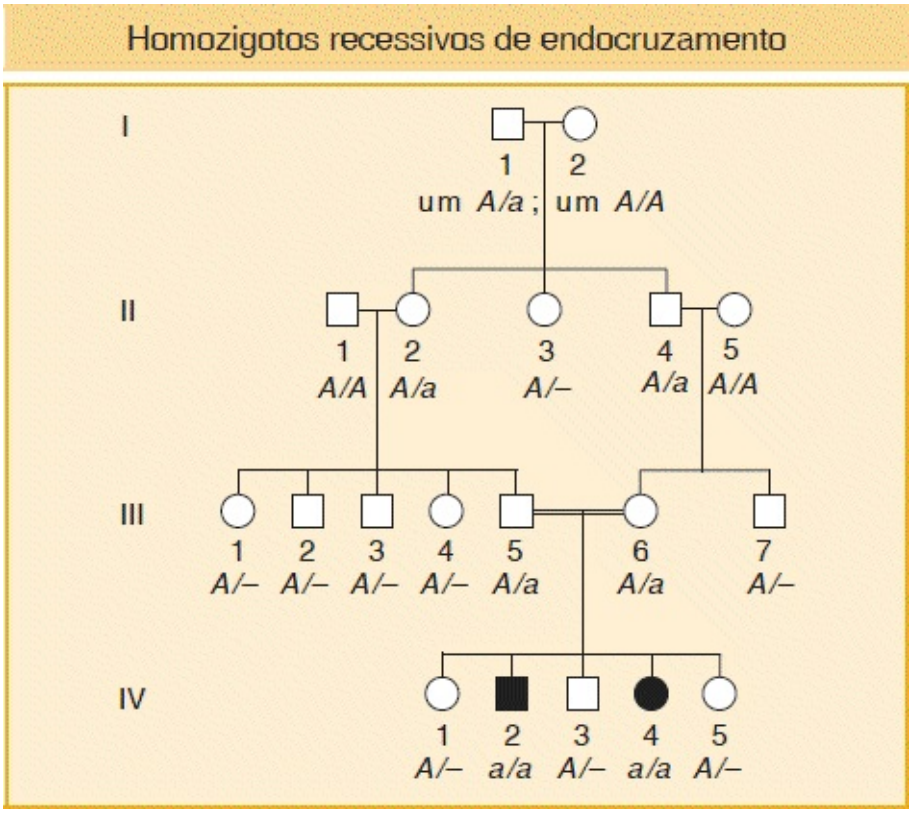


FIGURA 2.19 Heredograma de um fenótipo recessivo raro determinado por um alelo recessivo *a*. Os símbolos gênicos normalmente não são incluídos no heredograma, mas os genótipos são inseridos aqui para referência. As pessoas II-1 e II-5 se casam dentro da família; presume-se que sejam normais, tendo em vista que a condição hereditária em exame é rara. Observe também que não é possível ter certeza sobre o genótipo

de algumas pessoas com fenótipo normal; as referidas pessoas são indicadas por *A*/–. Os indivíduos III-5 e III-6, que geram os recessivos na geração IV, são primos em primeiro grau. Ambos herdaram seu alelo recessivo de um dos avós, seja I-1 ou I-2.

Os heredogramas familiares de distúrbios autossômicos recessivos tendem a aparentar ser um tanto quanto simples, com poucos símbolos pretos. Uma condição recessiva aparece em grupos de irmãos afetados e as pessoas em gerações anteriores e posteriores tendem a não ser afetadas. Para compreender por que isso é assim, é importante ter alguma compreensão sobre a estrutura genética das populações subjacentes às referidas condições raras. Por definição, se a condição for rara, a maior parte das pessoas não carreia o alelo anormal. Além disso, a maior parte daquelas pessoas que carreia o alelo anormal é heterozigota para ele, em vez de homozigota. O motivo básico de os heterozigotos serem muito mais comuns do que os homozigotos recessivos é que, para ser um homozigoto recessivo, ambos os genitores precisam apresentar o alelo *a*, mas, para ser um heterozigoto, apenas um genitor precisa apresentá-lo.

O nascimento de uma pessoa afetada normalmente depende da chance rara de união de genitores heterozigotos não aparentados. Entretanto, a endogamia (cruzamento entre parentes, por vezes denominado *consanguinidade* em seres humanos) aumenta a chance de que dois heterozigotos cruzem. Um exemplo de um casamento entre primos é demonstrado na Figura 2.19. Os indivíduos III-5 e III-6 são primos em primeiro grau e produzem dois homozigotos em relação ao alelo raro. Você pode observar a partir da Figura 2.19 que um ancestral que é um heterozigoto pode produzir muitos descendentes que também são heterozigotos. Portanto, dois primos podem carrear o *mesmo* alelo recessivo raro herdado de um ancestral em comum. Para que duas pessoas *não aparentadas* sejam heterozigotas, elas teriam herdado o alelo raro de *ambas* as suas famílias. Portanto, cruzamentos entre parentes em geral implicam um risco mais alto de produzir distúrbios recessivos do que cruzamentos entre não parentes. Por esse motivo, casamentos de primos em primeiro grau contribuem para uma grande proporção das pessoas com doenças recessivas na população.

Alguns outros exemplos de distúrbios recessivos humanos estão demonstrados na Figura 2.20. A fibrose cística é uma doença herdada no cromossomo 7 de

acordo com as regras mendelianas como um fenótipo autossômico recessivo. Seu sintoma mais importante é a secreção de grandes quantidades de muco para os pulmões, que resulta em morte em virtude de uma combinação de efeitos, mas normalmente precipitada pela infecção do sistema respiratório. O muco pode ser deslocado por tapotagem e a infecção pulmonar pode ser prevenida com antibióticos; portanto, com tratamento, pacientes com fibrose cística podem viver até a fase adulta. O gene da fibrose cística (e seu alelo mutante) foi um dos primeiros genes de doença humana a serem isolados no nível do DNA, em 1989. Essa linha de pesquisa finalmente revelou que o distúrbio é causado por uma proteína defeituosa que normalmente transporta íons cloreto pela membrana celular. A alteração resultante do equilíbrio eletrolítico altera a constituição do muco pulmonar. Essa nova compreensão sobre a função do gene em pessoas afetadas e não afetadas deu esperanças para um tratamento mais efetivo.

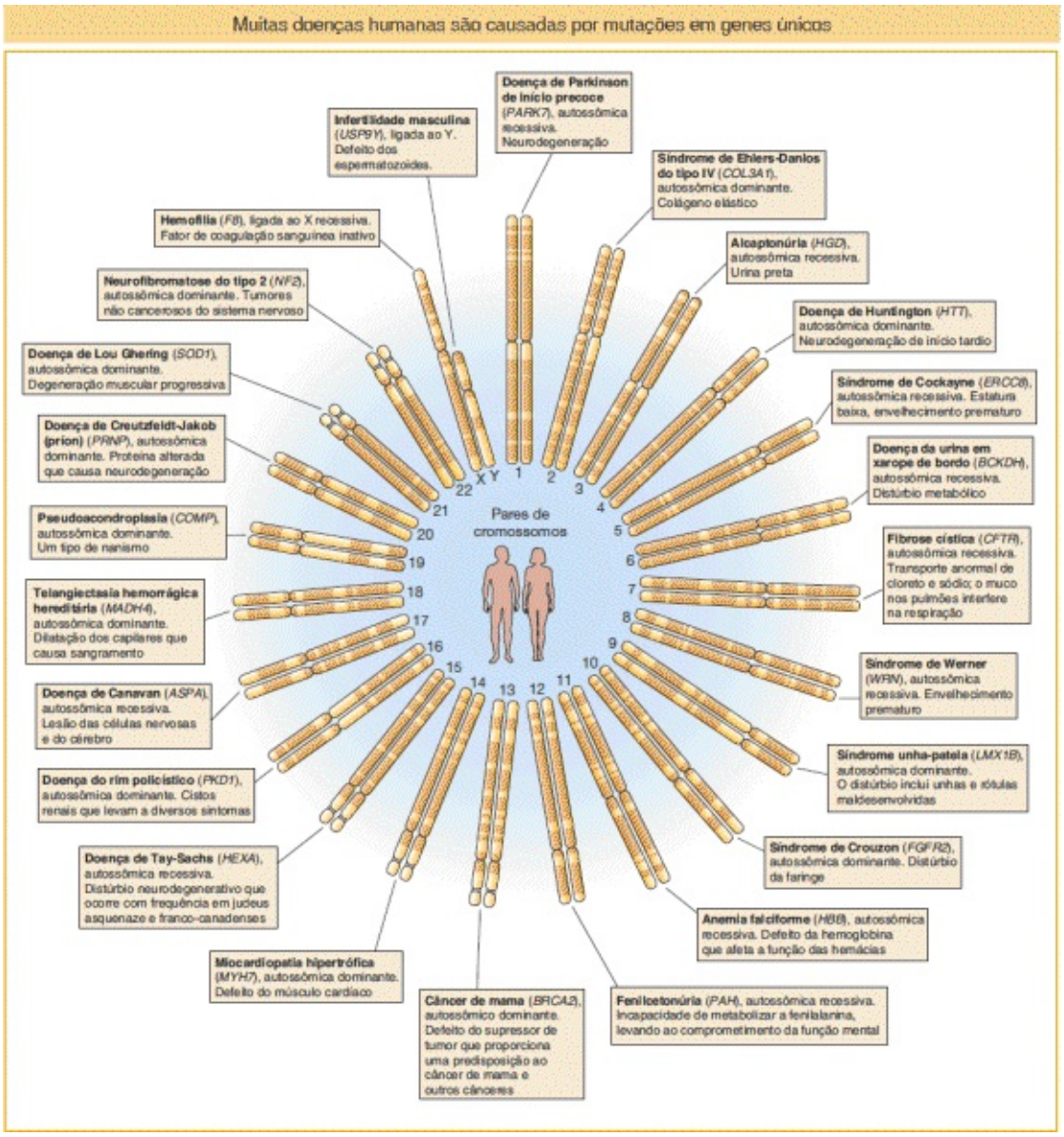


FIGURA 2.20 As posições dos genes mutados em algumas doenças monogênicas, mostradas nos 23 pares de cromossomos em um ser humano. Cada cromossomo apresenta um padrão de bandas característico. X e Y são os cromossomos sexuais (XX em mulheres e XY em homens). Os genes associados a cada doença estão demonstrados em parênteses.

O albinismo humano também é herdado de modo autossômico recessivo-padrão. O alelo mutante é de um gene que normalmente sintetiza o pigmento melanina marrom ou preto, normalmente encontrado na pele, nos cabelos e na retina do olho (Figura 2.21).

> CONCEITO-CHAVE Em heredogramas humanos, um distúrbio autossômico recessivo em geral é revelado pelo aparecimento do distúrbio na progênie masculina e feminina de pais não afetados.

Distúrbios autossômicos dominantes

Quais padrões de heredograma são esperados de distúrbios autossômicos dominantes? Aqui, o alelo normal é recessivo e o alelo defeituoso é dominante. Pode parecer um paradoxo que um distúrbio raro possa ser dominante, mas lembre que a dominância e a recessividade são simplesmente propriedades de como os alelos atuam em heterozigotos e não são definidas em referência a quão comuns eles são na população. Um bom exemplo de um fenótipo dominante raro que demonstra herança monogênica é a pseudoacondroplasia, um tipo de nanismo (Figura 2.22). Em relação a esse gene, pessoas com estatura normal são genotipicamente *d*/*d* e o fenótipo anão em princípio poderia ser *D*/d ou *D*/*D*. Entretanto, acredita-se que as duas "doses" do alelo *D* no genótipo *D*/*D* produzam um efeito tão grave que esse genótipo é letal. Se essa crença em geral for verdadeira, todos os anões são heterozigotos.

Na análise de heredogramas, os principais indícios para a identificação de um distúrbio autossômico dominante com herança mendeliana são o fato de o fenótipo tender a aparecer em todas as gerações do heredograma e de os pais ou mães afetados transmitirem o fenótipo para ambos os filhos e filhas. Novamente, a representação igual de ambos os sexos entre a progênie afetada descarta a herança por meio dos cromossomos sexuais. O fenótipo aparece em todas as gerações porque, em geral, o alelo anormal carreado por uma pessoa necessariamente provém de um genitor na geração precedente. (Alelos anormais também podem surgir *de novo* por mutações. Essa possibilidade tem de ser mantida em mente em relação a distúrbios que interferem com a reprodução, pois, aqui, a condição provavelmente não foi herdada de um genitor afetado.) Um heredograma típico de um distúrbio dominante está demonstrado na Figura 2.23. Mais uma vez, observe que as proporções mendelianas não são necessariamente observadas em famílias. Assim como nos distúrbios recessivos, pessoas que

contêm uma cópia do alelo *A* raro (*A*/*a*) são muito mais comuns do que aquelas que contêm duas cópias (*A*/*A*); assim, a maior parte das pessoas afetadas é heterozigota e virtualmente todos os cruzamentos que produzem progênie com distúrbios dominantes são *A*/*a* × *a*/*a*. Portanto, se a progênie dos referidos cruzamentos for totalizada, é esperada uma proporção de 1:1 de pessoas não afetadas (*a*/*a*) e afetadas (*A*/*a*).

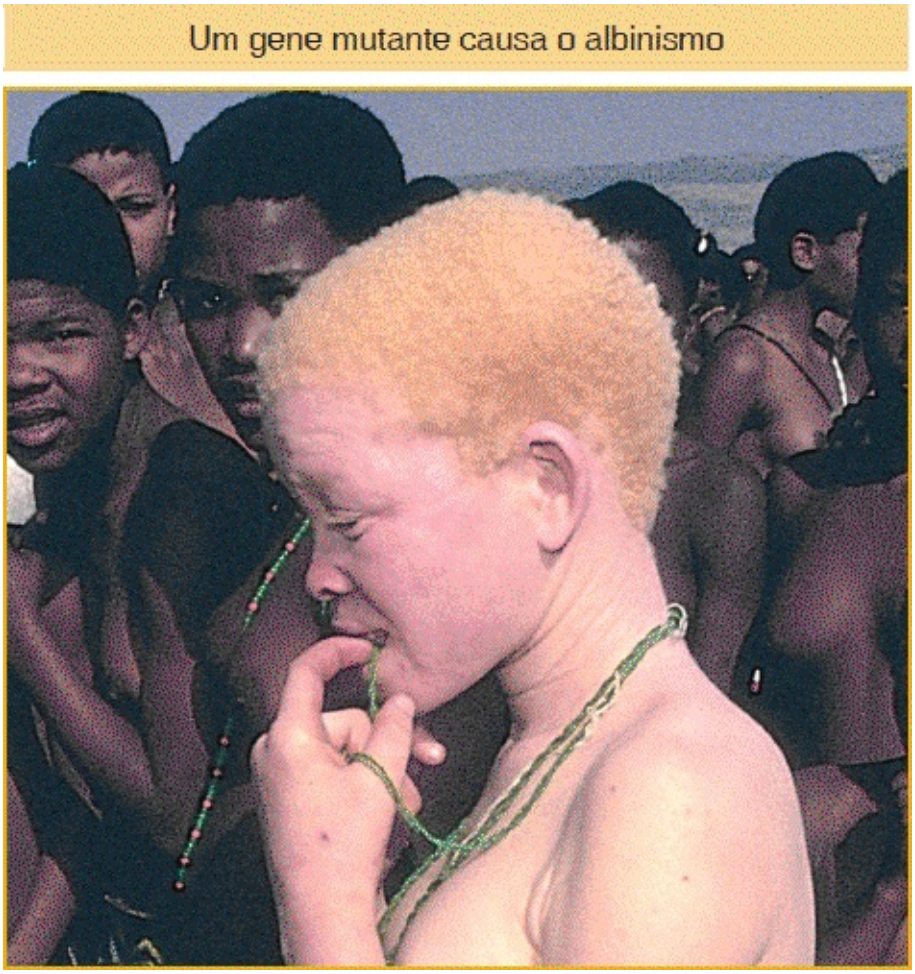


FIGURA 2.21 Uma versão não funcional de um gene que produz o pigmento da pele resulta em ausência de pigmento. Nesse caso, ambas as cópias do par de genes são mutadas. (Yves GELLIE/Gamma-Rapho/Getty Images.)

A doença de Huntington é um exemplo de uma doença herdada como um fenótipo dominante determinado por um alelo de um gene único. O fenótipo é de degeneração neural, que leva a convulsões e à morte prematura. O cantor de *folk*

Woody Guthrie sofria de doença de Huntington. A doença é um tanto quanto incomum, porque apresenta início tardio, os sintomas em geral não aparecem até que a pessoa tenha procriado (Figura 2.24). Quando a doença é diagnosticada em um dos pais, sabe-se que toda criança já nascida apresenta uma chance de 50% de herdar o alelo e a doença correlata. Esse padrão trágico inspirou um grande esforço para encontrar modos de identificar pessoas que carreiam o alelo anormal antes de manifestarem a doença. Atualmente existem exames moleculares para a identificação de pessoas portadoras do alelo da doença de Huntington.

Fenótipo de pseudoacondroplasia

FIGURA 2.22 O fenótipo da pseudoacondroplasia humana está ilustrado aqui por uma família de cinco irmãs e dois irmãos. O fenótipo é determinado por um alelo dominante, que podemos denominar *D*, que interfere no crescimento dos ossos longos durante o desenvolvimento. Esta fotografia foi feita quando a família chegou a Israel após o término da Segunda Guerra Mundial. (Bettmann/CORBIS.)

Algumas outras condições dominantes raras são a polidactilia (dedos extras), demonstrada na Figura 2.25, e o piebaldismo, demonstrado na Figura 2.26.

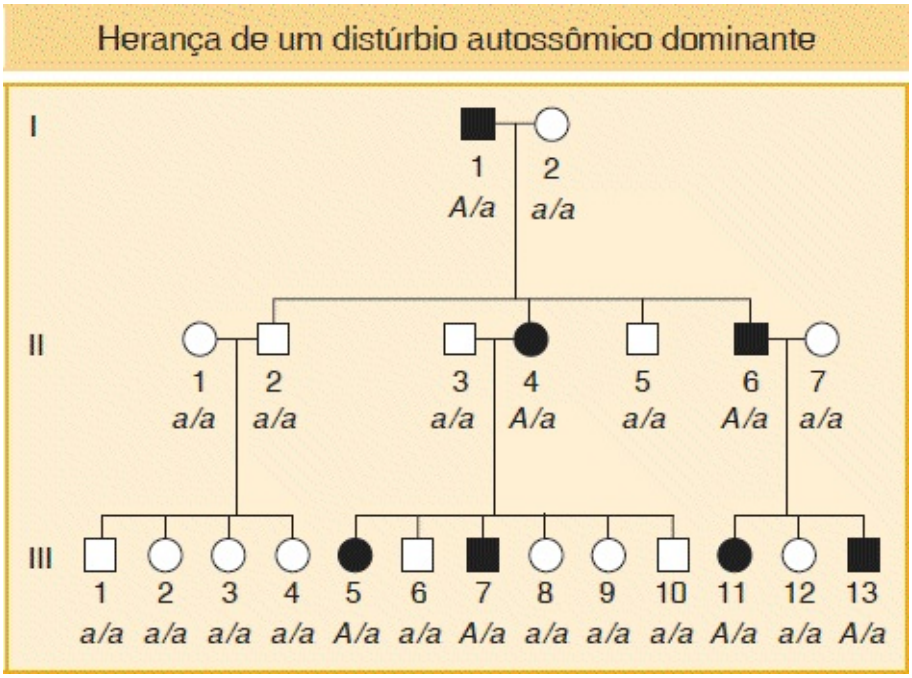


FIGURA 2.23 Heredograma de um fenótipo dominante determinado por um alelo *A* dominante. Neste heredograma, todos os genótipos foram deduzidos.

> CONCEITO-CHAVE Heredogramas de distúrbios mendelianos autossômicos dominantes demonstram indivíduos dos sexos masculino e feminino afetados em todas as gerações; eles também demonstram homens e mulheres afetados que transmitem a condição para proporções iguais de seus filhos e suas filhas.

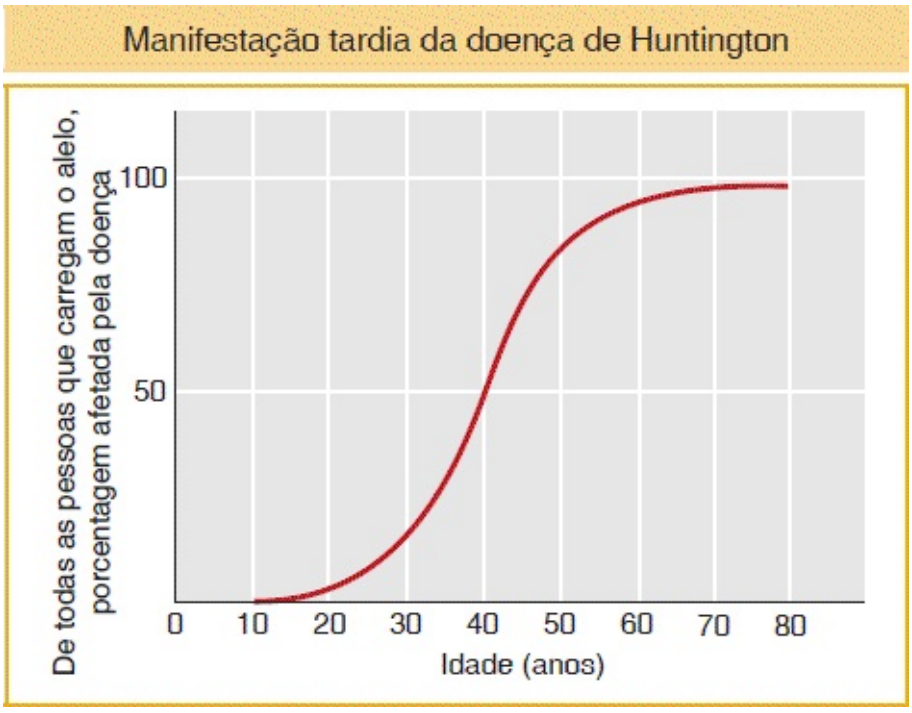


FIGURA 2.24 O gráfico demonstra que pessoas que carreiam o alelo em geral manifestam a doença após a idade fértil.

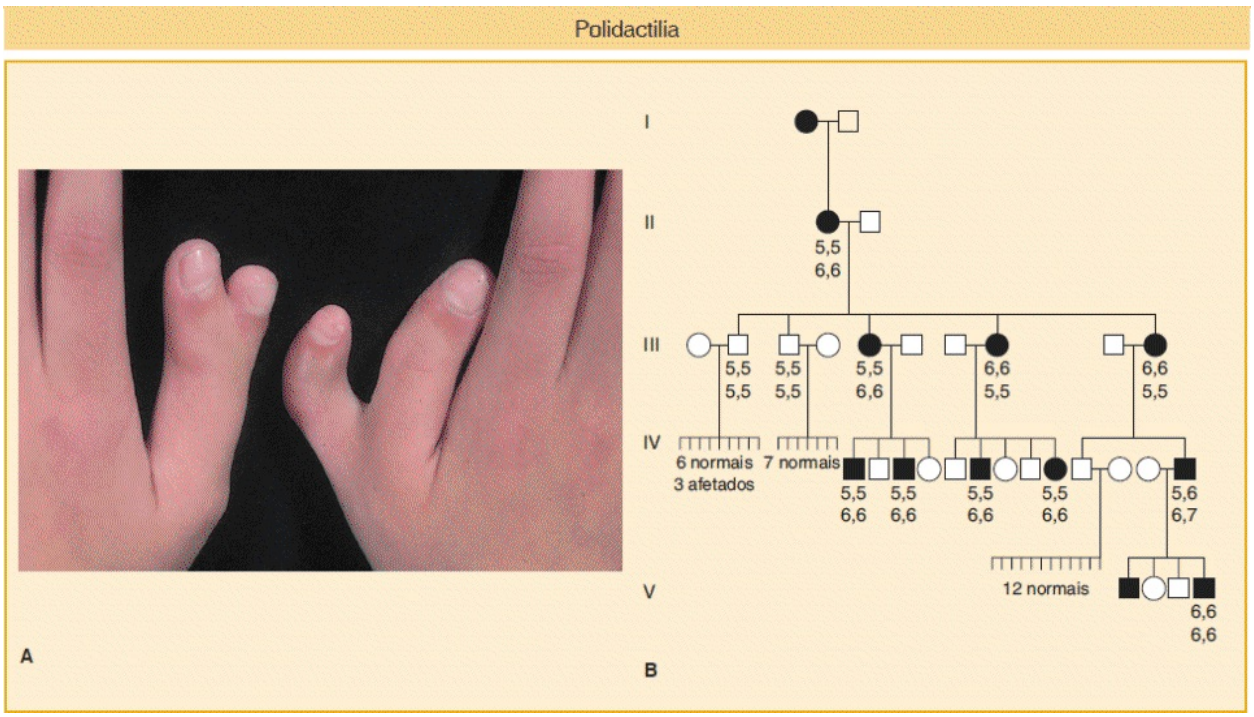


FIGURA 2.25 A polidactilia é um fenótipo dominante raro das mãos e dos pés humanos. A. A polidactilia, caracterizada por dedos extras das mãos ou dos pés, ou de ambos, é determinada por um alelo *P*. Os números

no heredograma (B) fornecem a quantidade de dedos das mãos nas linhas superiores e a quantidade de dedos dos pés nas linhas inferiores. (Observe a variação na expressão do alelo *P*). (*A.* ©Biophoto Associates/Science Source.)

Polimorfismos autossômicos

Os fenótipos alternativos de um polimorfismo (os **morfos**) com frequência são herdados como alelos de um gene autossômico único pelo modo mendeliano padrão. Entre os muitos exemplos humanos estão os **dimorfismos** (com dois morfos, os polimorfismos mais simples) a seguir: olhos castanhos *versus* azuis, cabelo pigmentado *versus* louro, capacidade de sentir o odor de frésias (um tipo de flor fragrante) *versus* incapacidade, pico de viúva *versus* nenhum, cerume viscoso *versus* seco e lóbulos da orelha presos *versus* livres. Em cada exemplo, o morfo determinado pelo alelo dominante é escrito primeiramente.

A interpretação dos heredogramas em relação aos polimorfismos é um pouco diferente daquela dos distúrbios raros, tendo em vista que, por definição, os morfos são comuns. Vejamos um heredograma em relação a um caso humano interessante. A maior parte das populações humanas é dimórfica em relação à capacidade de degustar a substância química feniltiocarbamida (PTC); ou seja, as pessoas conseguem detectá-la como um gosto ruim e amargo ou — para a grande surpresa e a descrença dos degustadores — podem não a sentir em absoluto. A partir do heredograma na Figura 2.27, podemos observar que dois degustadores por vezes produzem filhos não degustadores, o que torna claro que o alelo que confere a capacidade de sentir o gosto é dominante e que o alelo para a não degustação é recessivo. Observe na Figura 2.27 que quase todas as pessoas que entraram para a família carreiam o alelo recessivo, seja em uma condição heterozigota ou homozigota. Portanto, tal heredograma difere daqueles de distúrbios recessivos raros, em relação aos quais a presunção convencional é de que todos os que entram para a família são homozigotos normais. Tendo em vista que ambos os alelos da PTC são comuns, não é surpresa que todos, com exceção de um dos familiares nesse heredograma, tenham se casado com pessoas com no mínimo uma cópia do alelo recessivo.

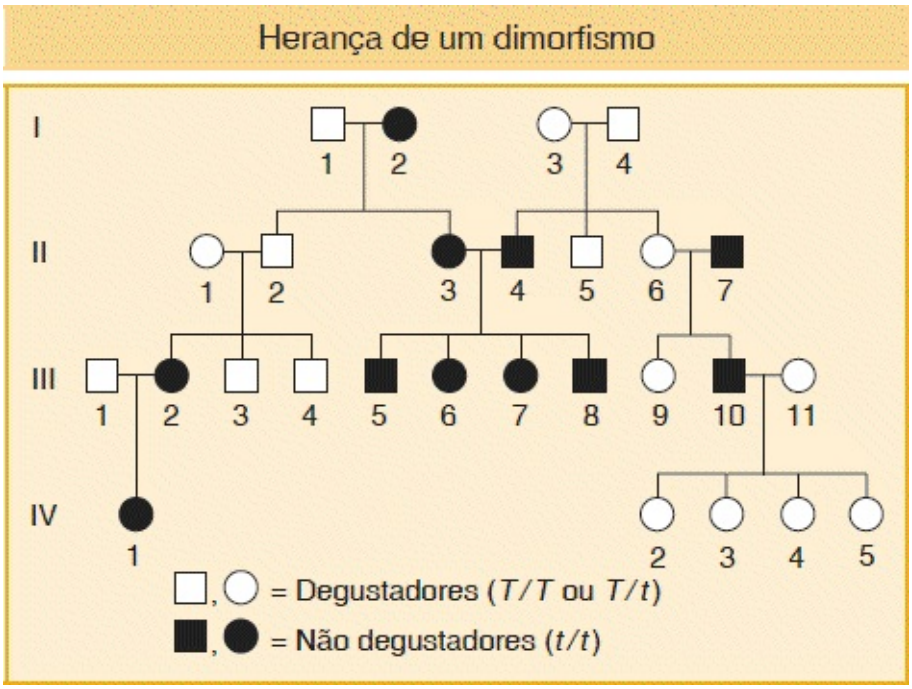


FIGURA 2.27 Heredograma em relação à capacidade de sentir o gosto da substância química feniltiocarbamida.

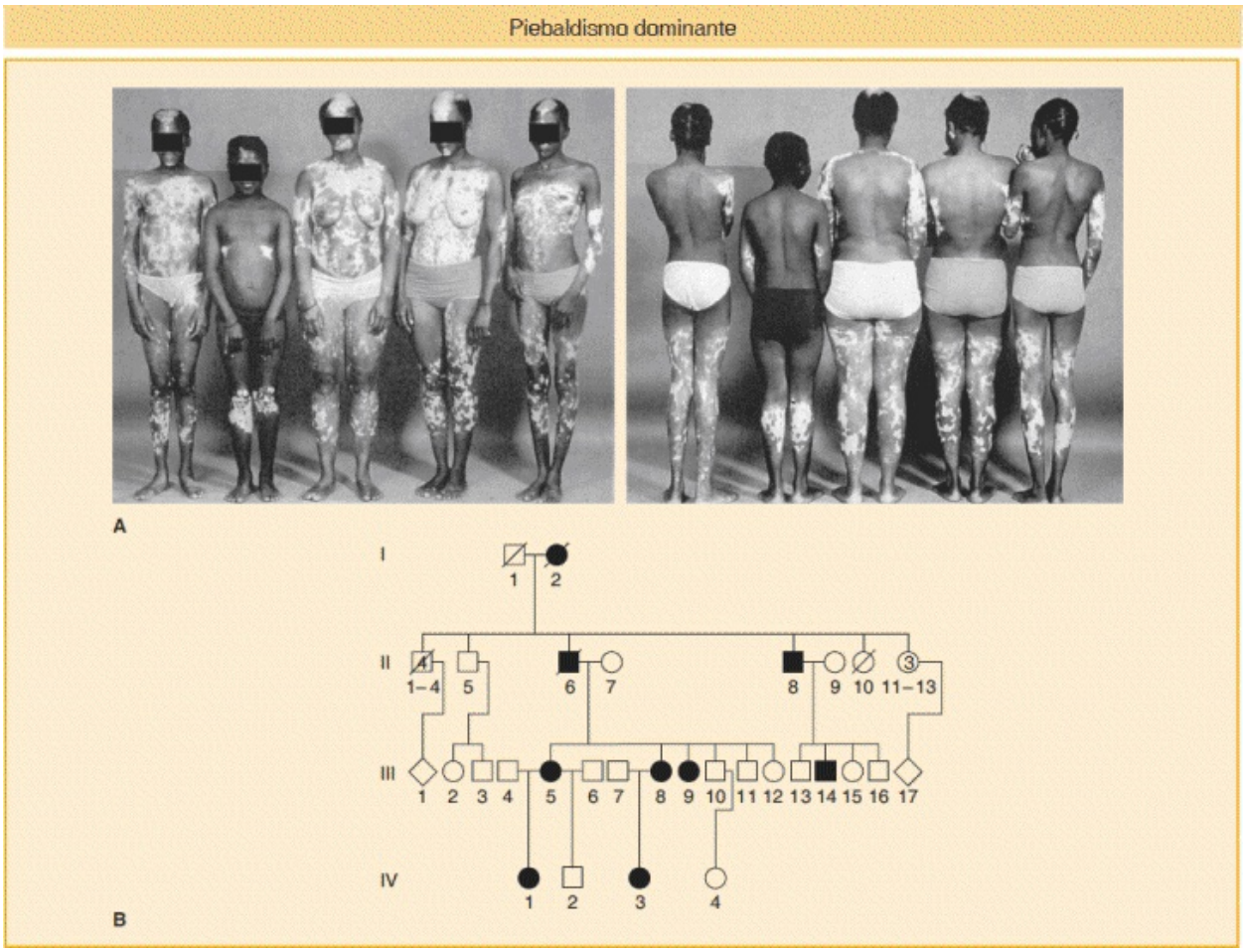


FIGURA 2.26 O piebaldismo é um fenótipo humano dominante raro. Embora o fenótipo seja encontrado esporadicamente em todas as raças, os padrões aparecem mais nos indivíduos com pele escura. As fotografias demonstram as vistas frontais e posteriores das pessoas afetadas IV-1 IV-3, III-5, III-8 e III-9 (**A**) do heredograma familiar (**B**). Observe a variação na expressão do gene do piebaldismo entre os familiares. Acredita-se que os padrões sejam causados pelo alelo dominante, que interfere na migração de melanócitos (células produtoras de melanina) da superfície dorsal para a ventral durante o desenvolvimento. A mancha branca na testa é particularmente característica e, com frequência, é acompanhada por uma mecha branca nos cabelos.

O piebaldismo não é um tipo de albinismo; as células nas áreas claras apresentam o potencial genético de produzir melanina, mas tendo em vista que não são melanócitos, não são programadas no seu desenvolvimento para fabricá-la. No albinismo verdadeiro, as células não apresentam o potencial de produzir melanina. (O piebaldismo é causado por mutações no *c-kit*, um tipo de gene denominado *proto-oncogene*; ver Capítulo 16.) (Fotos [*A*] e dados [*B*] de I. Winship, K. Young, R. Martell, R. Ramesar, D. Curtis e P. Beighton, "Piebaldism: An Autonomous Autosomal Dominant Entity", *Clin. Genet.* 39, 1991, 330. ©Reproduzida, com autorização, de John Wiley & Sons, Inc.)

O polimorfismo é um fenômeno genético interessante. Geneticistas de

populações surpreenderam-se ao descobrir quantos polimorfismos existem em populações naturais de plantas e animais em geral. Além disso, embora a genética de polimorfismos seja direta, existem muito poucos polimorfismos em relação aos quais existem explicações satisfatórias para a coexistência dos morfos. Mas os polimorfismos são abundantes em todos os níveis da análise genética, até mesmo no nível do DNA. De fato, os polimorfismos observados no nível do DNA têm sido inestimáveis como pontos de referência para ajudar os geneticistas a encontrar seu caminho nos cromossomos de organismos complexos, conforme será descrito no Capítulo 4. A genética de populações e evolutiva de polimorfismos é considerada nos Capítulos 17 e 19.

> CONCEITO-CHAVE As populações de plantas e animais (incluindo seres humanos) são altamente polimórficas. Morfos contrastantes com frequência são herdados como alelos de um gene único.

Distúrbios recessivos ligados ao X

Vejamos os heredogramas de distúrbios causados por alelos recessivos raros de genes localizados no cromossomo X. Os referidos heredogramas demonstram tipicamente as características a seguir:

1. Muito mais indivíduos do sexo masculino do que feminino apresentam o fenótipo raro em estudo. O motivo é que um indivíduo do sexo feminino só consegue herdar o genótipo se a sua mãe *e* o seu pai possuírem o alelo (p. ex., $X^A\ X^a \times X^a\ Y$), enquanto um indivíduo do sexo masculino pode herdar o fenótipo quando *apenas* a mãe carrear o alelo ($X^A\ X^a \times X^A\ Y$). Se o alelo recessivo for muito raro, quase todas as pessoas que apresentam o fenótipo são do sexo masculino.
2. Ninguém na progênie de um indivíduo do sexo masculino afetado apresenta o fenótipo, mas todas as suas filhas são “portadoras”, carreando o alelo recessivo mascarado na condição heterozigota. Na próxima geração, metade dos filhos dessas filhas portadoras apresenta o fenótipo (Figura 2.28).

3. Nenhum dos filhos homens de um indivíduo do sexo masculino afetado apresenta o fenótipo em estudo, nem transmitirá a condição para seus descendentes. O motivo por trás dessa ausência de transmissão entre indivíduos do sexo masculino é que um filho recebe seu cromossomo Y do pai; assim, normalmente ele não pode herdar o cromossomo X do pai também. Por outro lado, a transmissão entre indivíduos do sexo masculino de um distúrbio é um diagnóstico útil em relação a uma condição herdada de modo autossômico.

Na análise de heredogramas de recessivos ligados ao X raros, presume-se que um indivíduo do sexo feminino normal de genótipo desconhecido seja homozigoto, exceto se houver evidências em contrário.

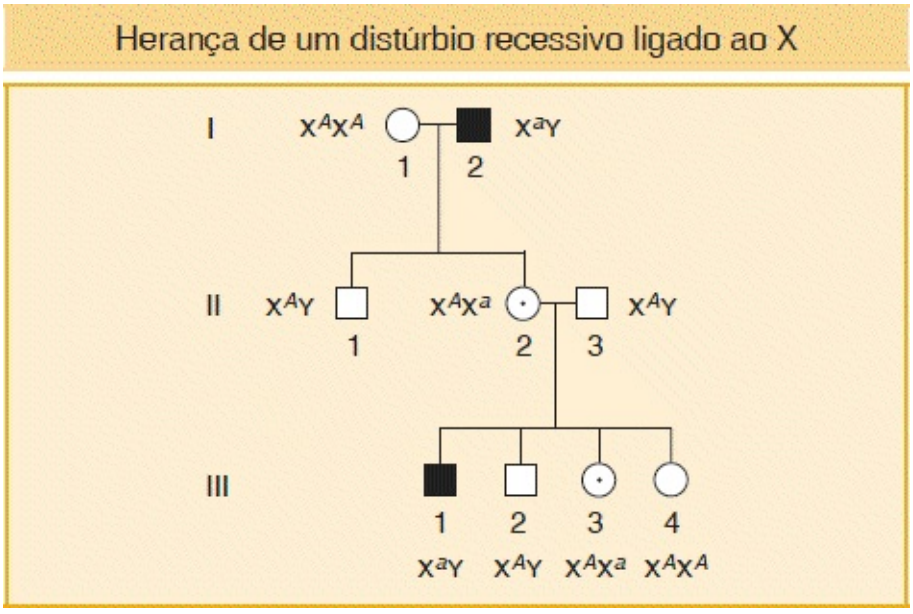


FIGURA 2.28 Conforme normalmente é o caso, a expressão dos alelos recessivos ligados ao X ocorre apenas em indivíduos do sexo masculino. Estes alelos são carreados, porém não expressos pelas filhas na próxima geração, para serem novamente expressos nos filhos. Observe que III-3 e III-4 não podem ser distinguidos fenotipicamente.

Talvez o exemplo mais familiar da herança recessiva ligada ao X seja o daltonismo. Pessoas com essa condição são incapazes de distinguir o vermelho do verde. Os genes para visão de cores foram caracterizados no nível molecular.

A visão de cores tem por base três diferentes tipos de cones na retina, cada um sensível aos comprimentos de onda vermelho, verde ou azul. Os determinantes genéticos em relação aos cones vermelhos e verdes encontram-se no cromossomo X. Pessoas daltônicas para vermelho e verde apresentam uma mutação em um desses dois genes. Assim como em qualquer distúrbio recessivo ligado ao X, existem muito mais homens com o fenótipo do que mulheres.

Outro exemplo familiar é a *hemofilia*, a falha do sangue em coagular. Muitas proteínas atuam em sequência para realizar a coagulação sanguínea. O tipo mais comum de hemofilia é causado pela ausência ou pela disfunção de uma dessas proteínas da coagulação, denominada *fator VIII*. Um heredograma de hemofilia bem conhecido é aquele das famílias reais inter-relacionadas da Europa (Figura 2.29). O alelo original da hemofilia no heredograma possivelmente surgiu espontaneamente como uma mutação nas células reprodutivas dos pais da Rainha Victoria ou na própria Rainha Victoria. Entretanto, alguns propuseram que a origem do alelo era um amante secreto da mãe de Victoria. Alexis, o filho do último czar da Rússia, herdou o alelo da hemofilia da Rainha Victoria, que era avó de sua mãe, Alexandra. Atualmente a hemofilia pode ser tratada clinicamente, mas anteriormente era uma condição potencialmente fatal. É interessante observar que o texto judaico do Talmude contém regras a respeito de isenções da circuncisão masculina, demonstrando claramente que o modo de transmissão da doença por meio de mulheres portadoras não afetadas era bem conhecido nos tempos antigos. Por exemplo, uma isenção era para filhos de mulheres cujos filhos das irmãs haviam sangrado profusamente quando foram circuncidados. Portanto, o sangramento anormal sabidamente era transmitido pelas mulheres da família, mas expresso apenas em filhos do sexo masculino.

A distrofia muscular de Duchenne é uma doença recessiva fatal ligada ao X. O fenótipo consiste em degeneração e atrofia dos músculos. Em geral, o início ocorre antes dos 6 anos de idade, com o confinamento a uma cadeira de rodas até os 12 anos e a morte até os 20. O gene para a distrofia muscular de Duchenne codifica a proteína muscular distrofina. Esse conhecimento implica esperanças para melhor compreensão sobre a fisiologia dessa condição e, finalmente, uma terapia.

Um fenótipo recessivo ligado ao X raro que é interessante do ponto de vista da diferenciação sexual é uma condição denominada *síndrome de feminização testicular*, que apresenta uma frequência de aproximadamente 1 em 65.000 nascimentos do sexo masculino. Pessoas com essa síndrome são cromossomicamente homens, apresentam 44 autossomos, mais um cromossomo X e um Y, mas se desenvolvem como indivíduos do sexo feminino (Figura 2.30). Eles apresentam genitália externa feminina, uma vagina com fundo cego e ausência de útero. Os testículos podem estar presentes nos lábios do pudendo ou no abdome. Embora muitas das referidas pessoas se casem, são estéreis. A condição não é revertida por meio do tratamento com o hormônio masculino androgênio e, assim, por vezes é denominada *síndrome da insensibilidade a andrógenos*. O motivo da insensibilidade é que uma mutação no gene do receptor de androgênio causa disfunção dos receptores e, assim, o hormônio masculino não tem efeito sobre os órgãos-alvo que contribuem para a masculinidade. Em humanos, a feminilidade resulta quando o sistema determinante da masculinidade não é funcional.

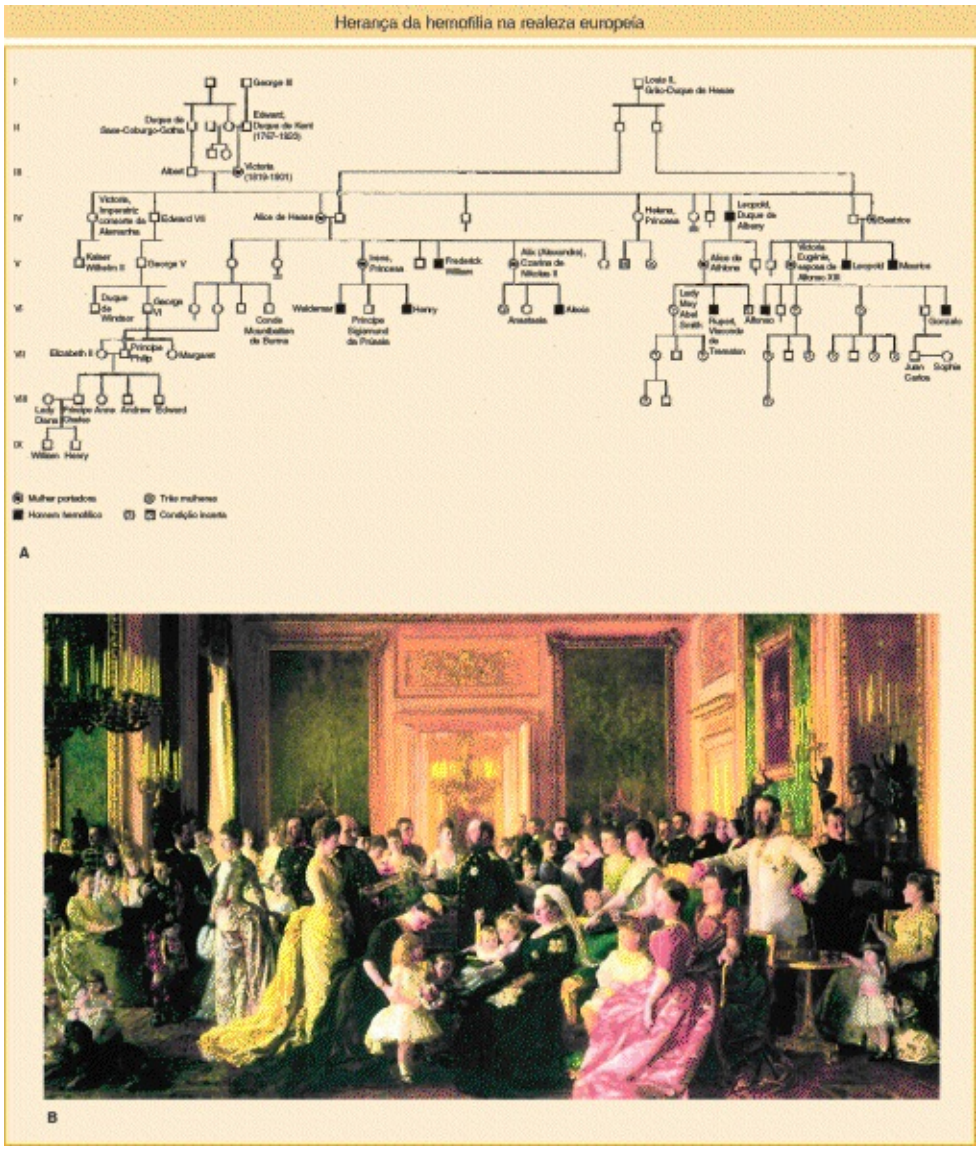


FIGURA 2.29 Heredograma da hemofilia, condição recessiva ligada ao X nas famílias reais da Europa. Um alelo recessivo que causa a hemofilia (comprometimento da coagulação sanguínea) surgiu por meio de uma mutação nas células reprodutivas da Rainha Victoria ou de um de seus pais. Esse alelo da hemofilia se propagou para outras famílias reais por meio do casamento entre elas. **A.** Esse heredograma parcial demonstra os homens afetados e as mulheres portadoras (heterozigotas). A maior parte das pessoas que se casaram com os familiares da Rainha Victoria foi omitida para simplificar o heredograma. Você consegue deduzir a probabilidade de a atual família real britânica possuir o alelo recessivo? **B.** Um quadro que demonstra a Rainha Victoria cercada por seus numerosos descendentes. (*B*. ©Lebrecht Music and Arts Photo Library/Alamy.)

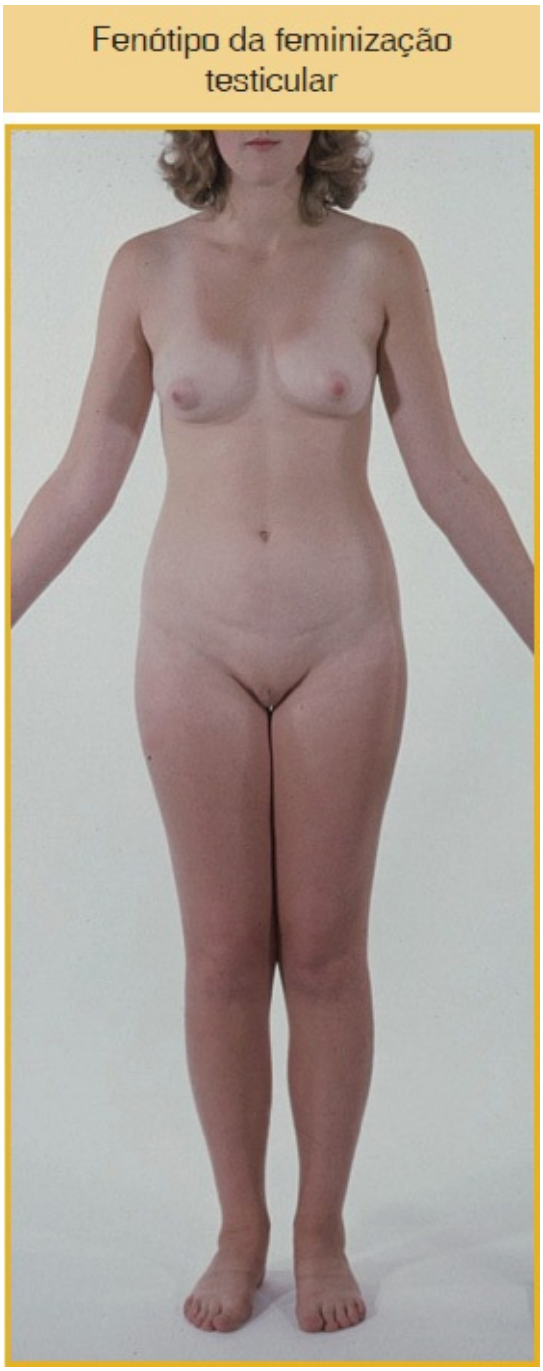


FIGURA 2.30 Um indivíduo XY com síndrome de feminização testicular, causada pelo alelo recessivo ligado ao X para a insensibilidade aos androgênios. (©Wellcome Photo Library/Custom Medical Stock.)

Distúrbios dominantes ligados ao X

Os padrões de herança dos distúrbios dominantes ligados ao X apresentam as características a seguir nos heredogramas (Figura 2.31):

1. Indivíduos do sexo masculino afetados transmitem a condição para todas as suas filhas, mas para nenhum de seus filhos.

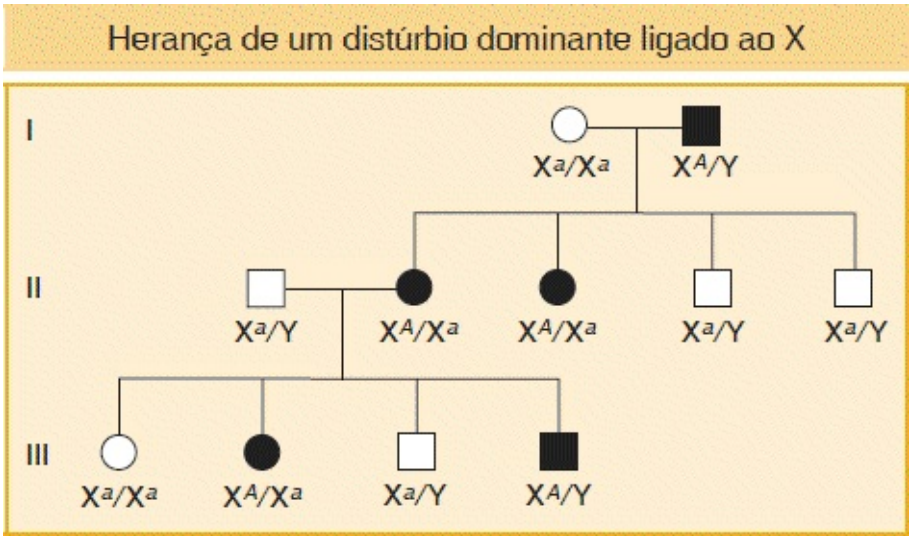


FIGURA 2.31 Todas as filhas de um indivíduo do sexo masculino que expressa um fenótipo dominante ligado ao X demonstrarão o fenótipo. Indivíduos do sexo feminino heterozigotos para um alelo dominante ligado ao X transmitirão a condição para metade de seus filhos e suas filhas.

2. Mulheres heterozigotas afetadas casadas com homens não afetados transmitem a condição para metade de seus filhos e de suas filhas.

Esse tipo de herança não é comum. Um exemplo é a hipofosfatemia, um tipo de raquitismo resistente à vitamina D. Alguns tipos de hipertricose (excesso de pelos corporais e faciais) demonstram herança dominante ligada ao X.

Herança ligada ao Y

Apenas os homens herdam genes localizados na região diferencial do cromossomo Y humano, com os pais transmitindo os genes aos seus filhos homens. O gene que desempenha um papel primário na masculinidade é o **gene *SRY***, por vezes denominado *fator determinante testicular*. A análise genômica confirmou que, de fato, o gene *SRY* encontra-se na região diferencial do cromossomo Y. Portanto, a própria masculinidade é ligada ao Y e demonstra o padrão esperado de transmissão exclusivamente de homem para homem. Demonstrou-se que alguns casos de esterilidade masculina são causados por

deleções de regiões do cromossomo Y que contêm genes promotores da espermatogênese. A esterilidade masculina não é hereditária, mas é interessante observar que os pais desses homens apresentam cromossomos Y normais, demonstrando que as deleções são novas.

Não houve casos convincentes de variantes fenotípicas não sexuais associadas ao cromossomo Y. Pelos nas bordas das orelhas (Figura 2.32) foram propostos como uma possibilidade, embora controversa. O fenótipo é extremamente raro entre as populações da maior parte dos países, mas é mais comum entre as populações da Índia. Em algumas famílias, os pelos nas bordas das orelhas são transmitidos exclusivamente do pai para os filhos.

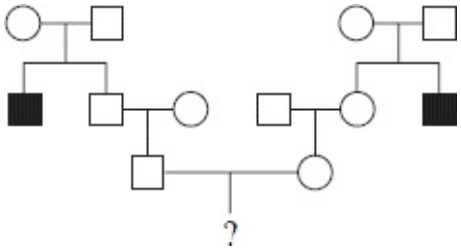
> CONCEITO-CHAVE Os padrões de herança com uma representação desigual dos fenótipos em indivíduos dos sexos masculino e feminino permitem localizar os genes envolvidos em um dos cromossomos sexuais.

FIGURA 2.32 Propôs-se que os pelos nas bordas das orelhas são causados por um alelo de um gene ligado ao Y (©Mark Collinson/Alamy.)

Cálculo de riscos na análise de heredogramas

Quando um distúrbio com uma herança monogênica bem-documentada sabidamente encontra-se presente em uma família, o conhecimento sobre os padrões de transmissão pode ser utilizado para calcular a probabilidade de os possíveis genitores terem um filho com o distúrbio. Por exemplo, considere um caso no qual marido e esposa recém-casados descobrem que cada um deles tinha um tio com doença de Tay-Sachs, uma doença autossômica recessiva grave, causada por disfunção da enzima hexosaminidase A. O defeito leva ao acúmulo

de depósitos adiposos nas células nervosas, causando paralisia seguida por morte precoce. O heredograma é como segue:



A probabilidade de o primeiro filho do casal apresentar doença de Tay-Sachs pode ser calculada da seguinte maneira. Tendo em vista que nenhum dos pais apresenta a doença, cada um somente pode ser um homozigoto normal ou um heterozigoto. Se ambos forem heterozigotos, então eles têm uma chance de transmitir o alelo recessivo para a criança, que então pode apresentar a doença de Tay-Sachs. Portanto, precisamos calcular a probabilidade de ambos serem heterozigotos e, assim sendo, a probabilidade de transmissão do alelo deletério para a criança.

1. Ambos os avós do marido eram necessariamente heterozigotos (*T*/*t*), tendo em vista que tiveram um filho *t*/*t* (o tio). Portanto, eles efetivamente constituíram um cruzamento mono-híbrido. O pai do marido poderia ser *T*/*T* ou *T*/*t*, mas dentro dos 3/4 da progênie não afetada sabemos que as probabilidades relativas desses genótipos devem ser de 1/4 e 1/2, respectivamente (a proporção da progênie esperada em um cruzamento mono-híbrido é $\frac{1}{4}$ *T*/*T*, $\frac{1}{2}$ *T*/*t*, $\frac{1}{4}$ *t*/*t*). Portanto, existe uma probabilidade de 2/3 de o pai ser um heterozigoto (dois terços é a proporção da progênie não afetada de heterozigotos: 2/4 divididos por 3/4).
2. Presume-se que a mãe do marido seja *T*/*T*, tendo em vista que ela entrou para a família e os alelos da doença em geral são raros. Portanto, se o pai for *T*/*t*, então o cruzamento com a mãe foi um cruzamento *T*/*t* × *T*/*T* e as proporções esperadas na progênie (que inclui o marido) são $\frac{1}{2}$ *T*/*T* e $\frac{1}{2}$ *T*/*t*.
3. A probabilidade geral de o marido ser um heterozigoto precisa ser calculada com a utilização de uma regra estatística denominada **regra do produto**, a

qual estabelece que:

> A probabilidade de dois eventos independentes ocorrerem é o produto de suas probabilidades individuais.

Tendo em vista que as transmissões de genes em diferentes gerações são eventos independentes, podemos calcular que a probabilidade de o marido ser um heterozigoto é a probabilidade de seu pai ser um heterozigoto (2/3) *multiplicada* pela probabilidade de seu pai ter um filho heterozigoto (1/2), que é 2/3 × 1/2 = 1/3.

4. De modo semelhante, a probabilidade de a esposa ser heterozigota também é de 1/3.
5. Se eles forem ambos heterozigotos (*T*/*t*), seu cruzamento será um cruzamento mono-híbrido padrão e, assim, a probabilidade de terem um filho *t*/*t* é de 1/4.
6. Em geral, a probabilidade do casal de ter um filho afetado é a probabilidade de ambos serem heterozigotos e em seguida de ambos transmitirem o alelo recessivo para um filho. Novamente, esses eventos são independentes e, assim, podemos calcular a probabilidade geral como 1/3 × 1/3 × 1/4 = 1/36. Em outras palavras, existe 1 chance em 36 de eles terem uma criança com doença de Tay-Sachs.

Em algumas comunidades judias, o alelo de Tay-Sachs não é tão raro como é na população geral. Nos referidos casos, não se pode presumir que pessoas não afetadas que se casam dentro de famílias com um histórico de Tay-Sachs sejam *T*/*T*. Se a frequência de heterozigotos *T*/*t* na comunidade for conhecida, essa frequência pode ser levada em conta no cálculo da regra do produto. Atualmente, estão disponíveis testes de diagnóstico molecular em relação aos alelos de Tay-Sachs e a utilização prudente desses testes reduziu drasticamente a frequência da doença em algumas comunidades.

RESUMO

Na divisão das células somáticas, o genoma é transmitido por meio da mitose, uma divisão nuclear. Nesse processo, cada cromossomo se replica em um par de

cromátides e as cromátides são separadas para produzir duas células-filhas idênticas. (A mitose pode ocorrer em células diploides ou haploides.) Na meiose, que ocorre no ciclo sexual em meiócitos, cada homólogo se replica para formar uma díade de cromátides; em seguida, as díades pareiam para formar uma tétrade, que sofre segregação em cada uma das duas divisões celulares. O resultado é quatro células haploides, ou gametas. A meiose pode ocorrer apenas em uma célula diploide; portanto, organismos haploides se unem para formar um meiócito diploide.

Um modo fácil de relembrar os principais eventos da meiose, por meio da utilização de seus dedos para representar os cromossomos, está demonstrado na Figura 2.33.

A dissecção genética de uma propriedade biológica tem início com uma coleção de mutantes. Cada mutante tem de ser testado para verificar se é herdado como uma alteração monogênica. O procedimento seguido mantém-se essencialmente inalterado desde a época de Mendel, que realizou a análise prototípica desse tipo. A análise tem por base a observação de proporções fenotípicas específicas na progênie de cruzamentos controlados. Em um caso típico, um cruzamento de *A*/*A* × *a*/*a* produz uma F_1 que é toda *A*/*a*. Quando a F_1 é autofecundada ou entrecruzada, uma proporção genotípica de $\frac{1}{4}$ *A*/*A* : $\frac{1}{2}$ *A*/*a* : $\frac{1}{4}$ *a*/*a* é produzida na F_2. (No nível fenotípico, essa proporção é de $\frac{3}{4}$ *A*/— : $\frac{1}{4}$ *a*/*a*.) Os três genótipos monogênicos são homozigoto dominante, heterozigoto (mono-híbrido) e homozigoto recessivo. Se um indivíduo *A*/*a* for cruzado com um *a*/*a* (um cruzamento-teste), é produzida uma proporção de 1:1 na progênie. As proporções de 1:1, 3:1 e 1:2:1 têm origem no princípio da segregação igual, isso é, os produtos haploides da meiose de *A*/*a* serão $\frac{1}{2}$ *A* e $\frac{1}{2}$ *a*. A base celular da segregação igual dos alelos é a segregação de cromossomos homólogos na meiose. Fungos haploides podem ser utilizados para demonstrar a segregação igual no nível de uma meiose única (uma proporção de 1:1 em um asco).

A base molecular para a produção das cromátides na meiose é a replicação do DNA. A segregação na meiose pode ser observada diretamente no nível molecular (DNA). A força molecular da segregação é a despolimerização e o

subsequente encurtamento dos microtúbulos que estão unidos aos centrômeros. As mutações recessivas em geral ocorrem em genes haplossuficientes, enquanto as mutações dominantes com frequência ocorrem em virtude da haploinsuficiência gênica.

Em muitos organismos, o sexo é determinado cromossomicamente e, tipicamente, XX é o sexo feminino e XY é o sexo masculino. Os genes no cromossomo X (genes ligados ao X) não apresentam correspondentes no cromossomo Y e demonstram um padrão de herança monogênica que difere nos dois sexos, com frequência resultando em proporções diferentes nas progênies masculina e feminina.

A segregação monogênica mendeliana é útil na identificação dos alelos mutantes subjacentes a muitos distúrbios humanos. A análise de heredogramas pode revelar distúrbios autossômicos ou ligados ao X de ambos os tipos, dominante e recessivo. A lógica da genética mendeliana precisa ser utilizada com cautela, levando em consideração que o tamanho da progênie humana é pequeno e as proporções fenotípicas não são necessariamente típicas daquelas esperadas em tamanhos de amostra maiores. Se um distúrbio monogênico conhecido estiver presente em um heredograma, a lógica mendeliana pode ser utilizada para prever a probabilidade de os filhos herdarem a doença.

Os principais eventos da mitose e da meiose

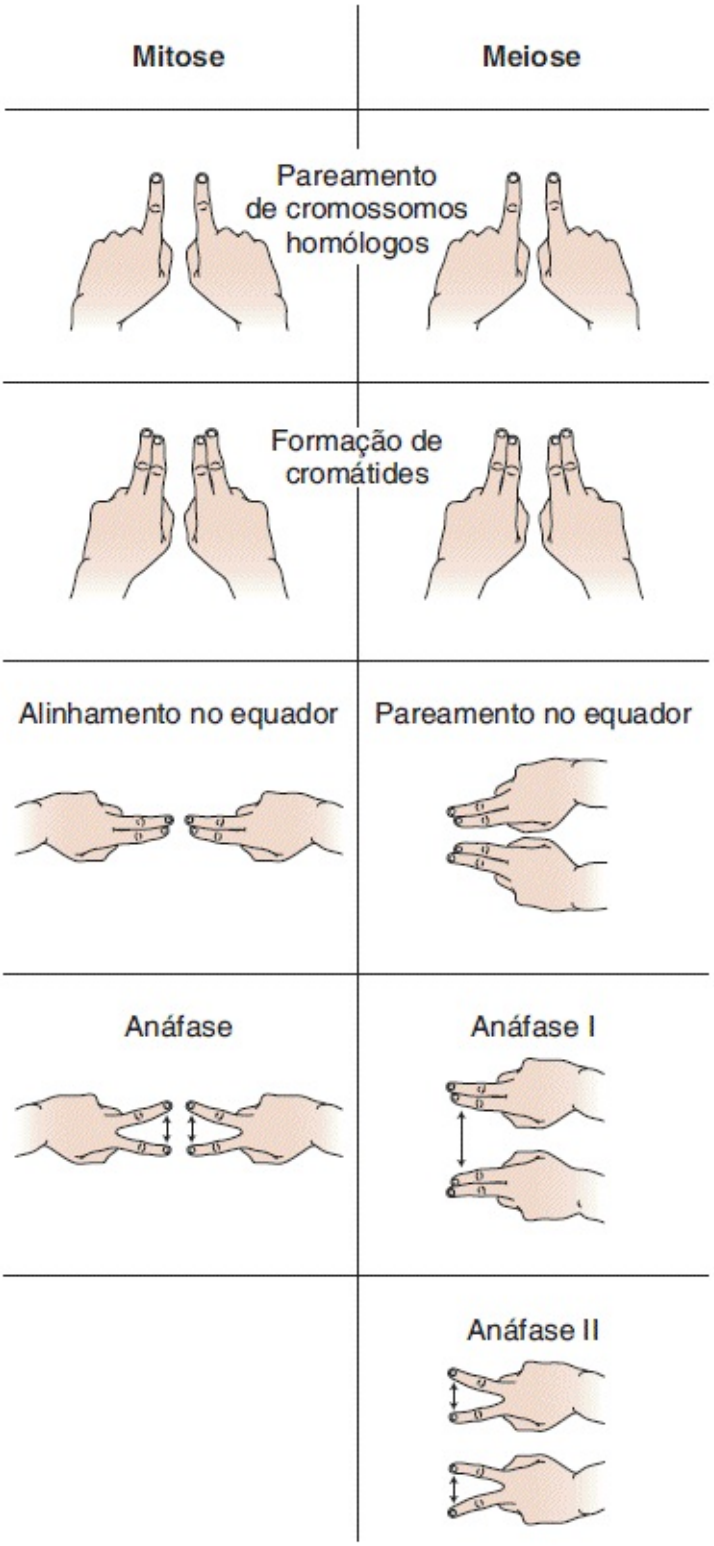


FIGURA 2.33 Utilização dos dedos das mãos para relembrar os principais eventos da mitose e meiose.

TERMOS-CHAVE

alelo
alelo nulo
análise de heredogramas
asco
autopolinização
bivalente
característica
cromátide
cromossomo sexual
cromossomo X
cromossomo Y
cruzamento
cruzamento mono-híbrido
cruzamento-teste
descoberta dos genes
díade
dimorfismo
dissecção genética
dominante
espécie dioica
fenótipo
gene
gene *SRY*
genética direta
genética reversa
genótipo
geração parental (P)
haploinsuficiente

haplossuficiente
hemizigoto
heterozigoto
heterozigose
homozigoto
homozigoto dominante
homozigoto recessivo
lei da segregação igual (primeira lei de Mendel)
ligação ao sexo
ligação ao X
ligação ao Y
linhagem pura
meiócito
meiose
mitose
mono-híbrido
morfo
mutação
mutação hipomórfica (*leaky*)
mutante
polimorfismo
primeira geração filial (F_1)
probando
produto da meiose
propriedade
recessivo
regiões pseudoautossômicas 1 e 2
regra do produto
segunda geração filial (F_2)
sexo heterogamético
sexo homogamético
testador

tétrade
tipo selvagem
traço
zigoto

PROBLEMAS RESOLVIDOS

Esta seção em cada capítulo contém alguns poucos problemas resolvidos, que demonstram como abordar os conjuntos de problemas que se seguem. A finalidade dos conjuntos de problemas é desafiar a sua compreensão sobre os princípios genéticos aprendidos no capítulo. O melhor modo de demonstrar a compreensão sobre um assunto é ser capaz de utilizar aquele conhecimento em uma situação real ou simulada. Esteja alerta que não existe um modo semelhante ao de uma máquina de solucionar esses problemas. As três fontes principais à sua disposição são os princípios genéticos aprendidos há pouco, a lógica e a tentativa e o erro.

Existem alguns conselhos gerais antes de iniciar. Primeiramente, é absolutamente essencial ler e compreender todo o problema. A maior parte dos problemas utiliza dados obtidos a partir de uma pesquisa que alguém realmente realizou: pergunte-se por que a pesquisa pode ter sido iniciada e qual era o objetivo provável. Descubra exatamente quais fatos são fornecidos, quais presunções precisam ser feitas, quais indicações são fornecidas no problema e quais inferências podem ser feitas a partir das informações disponíveis. Em segundo lugar, seja metódico. Contemplar o problema raramente é útil. Reformule as informações no problema da sua própria maneira, preferencialmente com a utilização de uma representação diagramática ou um fluxograma para lhe ajudar a pensar sobre o problema. Boa sorte.

Problema resolvido 1. Foram realizados cruzamentos entre duas linhagens puras de coelhos, que podemos denominar A e B. Um macho da linhagem A foi cruzado com uma fêmea da linhagem B e os coelhos da F_1 foram subsequentemente intercruzados para produzir uma F_2. Descobriu-se que três quartos dos animais da F_2 apresentavam gordura subcutânea branca e um quarto apresentava gordura

subcutânea amarela. Posteriormente, a F_1 foi examinada e observou-se que apresentava gordura branca. Diversos anos depois, foi realizada uma tentativa de repetir o experimento com a utilização do mesmo macho da linhagem A e a mesma fêmea da linhagem B. Nessa ocasião, a F_1 e a F_2 (22 animais) apresentavam gordura branca. A única diferença entre o experimento original e a repetição que aparentou ser relevante foi que, no original, todos os animais foram alimentados com vegetais frescos, enquanto na repetição foram alimentados com ração comercial para coelhos. Forneça uma explicação para a diferença e um teste da sua ideia.

Solução

Na primeira vez que o experimento foi realizado, os criadores teriam justificado perfeitamente a proposta de que um par de alelos determina a gordura corporal branca *versus* amarela, tendo em vista que os dados se assemelham claramente aos resultados de Mendel nas ervilhas. O branco tem de ser dominante, de modo que podemos representar o alelo branco como W e o alelo amarelo como *w*. Os resultados então podem ser expressos como segue:

P	$W/W \times w/w$
F_1	W/w
F_2	$\frac{1}{4}\ W/w$
	$\frac{1}{2}\ W/w$
	$\frac{1}{4}\ w/w$

Sem dúvidas, se os coelhos parentais houvessem sido sacrificados, se poderia prever que um genitor (não podemos dizer qual) apresentava gordura branca e o outro, amarela. Felizmente, os coelhos não foram sacrificados e os mesmos animais foram cruzados novamente, levando a um resultado muito interessante e diferente. Com frequência na ciência, uma observação inesperada pode levar a um novo princípio e, em vez de prosseguir para algo diferente, é útil tentar

explicar a inconsistência. Assim, por que a proporção de 3:1 desapareceu? Aqui estão algumas possíveis explicações.

Primeiramente, talvez os genótipos dos animais parentais houvessem alterado. Esse tipo de alteração espontânea que afeta todo o animal, ou no mínimo as suas gônadas, é muito improvável, tendo em vista que até mesmo a experiência comum nos informa que os organismos tendem a ser estáveis em seu tipo.

Em segundo lugar, na repetição, a amostra de 22 animais da F_2 não continha qualquer gordura amarela simplesmente ao acaso ("chance"). Novamente, essa explicação aparenta ser improvável, tendo em vista que a amostra era consideravelmente grande, mas essa é uma possibilidade.

Uma terceira explicação advém do princípio de que os genes não atuam em um vácuo; eles dependem do ambiente para os seus efeitos. Portanto, a fórmula "Genótipo + Ambiente = Fenótipo" é um mnemônico útil. Um corolário dessa fórmula é que os genes podem atuar de modo diferente em diferentes ambientes; assim,

Genótipo 1 + Ambiente 1 = Fenótipo 1

e

Genótipo 1 + Ambiente 2 = Fenótipo 2

No presente problema, as dietas diferentes constituíram os ambientes diferentes e, assim, uma explicação possível dos resultados é que o homozigoto recessivo *w*/*w* produz gordura amarela apenas quando a dieta contém vegetais frescos. Essa explicação pode ser testada. Um modo de testá-la é repetir o experimento novamente e utilizar vegetais frescos como alimento, mas os genitores podem estar mortos nessa ocasião. Um modo mais convincente é cruzar diversos coelhos da F_2 com gordura branca do segundo experimento. De acordo com a interpretação original, alguns deles devem ser heterozigotos e, se a sua progênie for criada com vegetais, a gordura amarela deve aparecer em proporções mendelianas. Por exemplo, se um cruzamento for *W*/*w* e *w*/*w*, a progênie será ½ com gordura branca e ½ com gordura amarela.

Se esse resultado não ocorrer e nenhum descendente com gordura amarela

aparecer em quaisquer dos cruzamentos, seremos forçados de volta à primeira ou à segunda explicação. A segunda explicação pode ser testada por meio da utilização de quantidades maiores e se essa explicação não funcionar, nos resta a primeira explicação, que é difícil de ser testada diretamente.

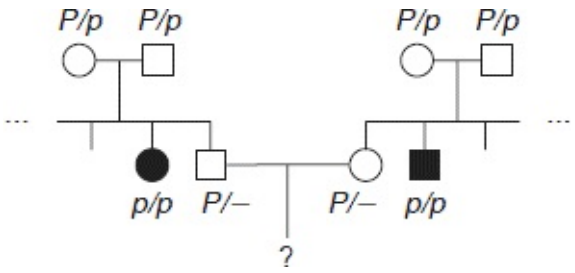
Como você poderia ter percebido, na realidade a dieta era a culpada. Os detalhes específicos ilustram muito bem os efeitos ambientais. Os vegetais frescos contêm substâncias amarelas denominadas xantofilas e o alelo dominante *W* proporciona aos coelhos a capacidade de fragmentar essas substâncias até um tipo incolor ("branco"). Entretanto, os animais *w*/*w* não têm essa capacidade e as xantofilas são depositadas na gordura, tornando-a amarela. Quando nenhuma xantofila tiver sido ingerida, ambos os animais *W*/— e *w*/*w* apresentarão gordura branca.

Problema resolvido 2. A fenilcetonúria (PKU) é uma doença hereditária humana que resulta da incapacidade do corpo de processar a substância química fenilalanina, que está contida nas proteínas que ingerimos. A PKU é manifestada no início da infância e, se permanecer sem tratamento, em geral leva ao retardo mental. A PKU é causada por um alelo recessivo com herança mendeliana simples.

Um casal pretende ter filhos, mas aconselha-se com um consultor genético, uma vez que o homem tem uma irmã com PKU e a mulher tem um irmão com PKU. Não existem outros casos conhecidos em suas famílias. Eles solicitam ao consultor genético que determine a probabilidade de que seu primeiro filho apresente PKU. Qual é a probabilidade?

Solução

O que podemos deduzir? Se denominarmos o alelo que causa o fenótipo da PKU de *p* e o respectivo alelo normal de *P*, então a irmã e o irmão do homem e da mulher, respectivamente, necessariamente são *p*/*p*. Para produzir essas pessoas afetadas, todos os quatro avós tinham de ser heterozigotos normais. O *pedigree* pode ser resumido como segue:



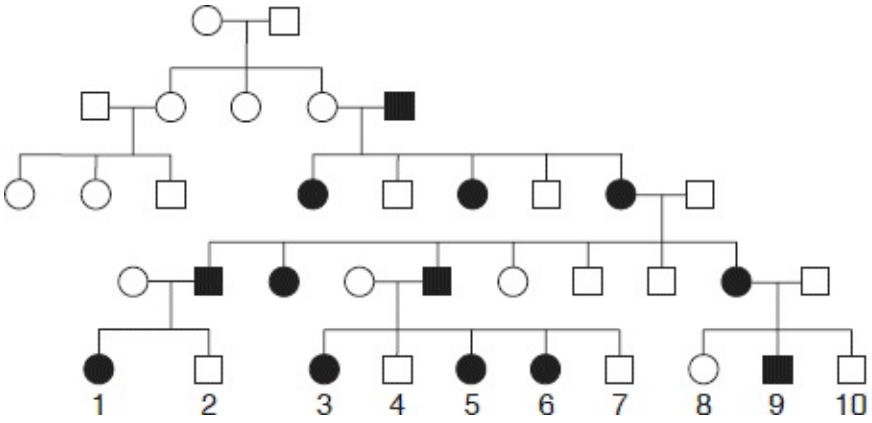
Quando essas inferências tiverem sido feitas, o problema é reduzido a uma aplicação da regra do produto. O único modo por meio do qual o homem e a mulher podem ter um filho com PKU é se ambos forem heterozigotos (é óbvio que eles próprios não apresentam a doença). Ambos os cruzamentos dos avós são cruzamentos mono-híbridos mendelianos simples, que se espera produzirem progênie nas proporções a seguir:

$$\left.\begin{array}{l}\frac{1}{4}\,P/P\\ \frac{1}{2}\,P/p\end{array}\right\}\ \text{Normal}\ (\tfrac{3}{4})$$
$$\frac{1}{4}\,p/p \quad \text{PKU}\ (\tfrac{1}{4})$$

Sabemos que o homem e a mulher são normais e, assim, a probabilidade de cada um ser um heterozigoto é de 2/3, tendo em vista que, dentro da classe *P*/—, 2/3 são *P*/*p* e 1/3 é *P*/*P*.

A probabilidade de *ambos* o homem e a mulher serem heterozigotos é 2/3 × 2/3 = 4/9. Se ambos forem heterozigotos, então um quarto de seus filhos apresentará PKU e, assim, a probabilidade de seu primeiro filho apresentar PKU é de 1/4 e a probabilidade de serem heterozigotos *e* de seu primeiro filho apresentar PKU é 4/9 × 1/4 = 4/36 = 1/9, que é a resposta.

Problema resolvido 3. Uma doença humana rara é observada em uma família, conforme demonstrado no heredograma a seguir.



a. Deduza o modo de herança mais provável.

b. Quais seriam os desfechos dos casamentos dos primos 1 × 9, 1 × 4, 2 × 3 e 2 × 8?

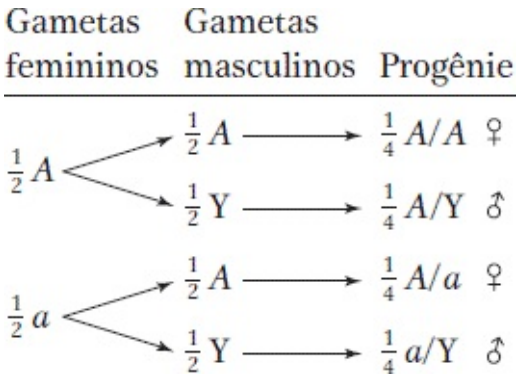
Solução

a. O modo mais provável de herança é o dominante ligado ao X. Presumimos que o fenótipo da doença seja dominante, tendo em vista que após ter sido introduzido no heredograma pelo homem na geração II, ele aparece em todas as gerações. Presumimos que o fenótipo seja ligado ao X porque não há transmissão paterna para os filhos homens. Se ele fosse autossômico dominante, a transmissão de pai para filho seria comum.

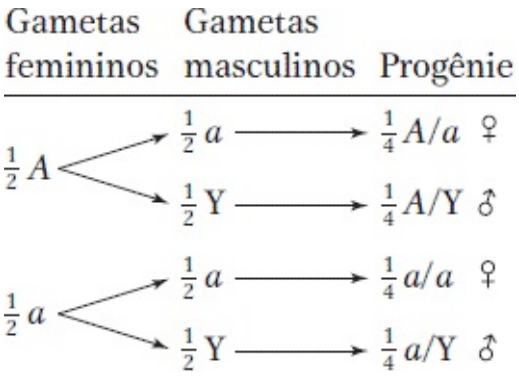
Em teoria, a herança autossômica recessiva poderia ocorrer, mas é improvável. Em particular, observe os casamentos entre os membros afetados da família e os estranhos não afetados. Se a condição fosse autossômica recessiva, o único modo por meio do qual esses casamentos poderiam ter descendência afetada seria se cada pessoa que entrou para a família fosse heterozigota; então os cruzamentos seriam *a*/*a* (afetado) × *A*/*a* (não afetado). Entretanto, nos é dito que a doença é rara; nesse caso, é altamente improvável que os heterozigotos sejam tão comuns. A herança recessiva ligada ao X é impossível, tendo em vista que um cruzamento de uma mulher afetada com um homem normal não produziria filhas afetadas. Assim, podemos deixar *A* representar o alelo que causa a doença e *a* representar o alelo normal.

b. 1 × 9: O número 1 tem de ser heterozigoto *A*/*a*, tendo em vista que ela

necessariamente obteve *a* de sua mãe normal. O número 9 deve ser *A*/Y. Portanto, o cruzamento é ♀ *A*/*a* × ♂ *A*/Y.



1 × 4: Tem de ser ♀ *A*/*a* × ♂ *a*/Y.



2 × 3: Tem de ser ♂ *a*/Y × ♀ *A*/*a* (o mesmo que 1 × 4).
2 × 8: Tem de ser ♂ *a*/Y × ♀ *a*/*a* (toda a progênie normal).

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

(As primeiras 13 questões requerem a inspeção das figuras do capítulo.)

1. Na parte à esquerda da Figura 2.4, as setas vermelhas demonstram autofecundação de flores únicas de uma planta da F_1. Os mesmos resultados da F_2 seriam produzidos por meio da polinização cruzada de duas plantas da F_1 diferentes?

2. Na parte à direta da Figura 2.4, na planta que demonstra uma proporção de

11:11, você acredita que seria possível encontrar uma vagem com todas as ervilhas amarelas? Todas verdes? Explique.

3. Na Tabela 2.1, declare o fenótipo recessivo em cada um dos sete casos.
4. Considerando a Figura 2.8, a sequência "pareamento → replicação → segregação → segregação" é uma boa descrição abreviada da meiose?
5. Aponte todos os casos de bivalentes, díades e tétrades na Figura 2.11.
6. Na Figura 2.11, presuma (como nas plantas do milho) que *A* codifica um alelo que produz amido no pólen e que o alelo *a* não produz. A solução de iodo cora o amido em preto. Como você demonstraria a primeira lei de Mendel diretamente com tal sistema?
7. Considerando a Figura 2.13, se você tivesse um mutante duplo homozigoto *m3*/*m3 m5*/*m5*, esperaria um fenótipo mutante? (Nota: essa linhagem apresentaria dois sítios mutantes na mesma sequência codificadora.)
8. Em quais dos estágios do ciclo de vida da *Drosophila* (representado no quadro Organismo-modelo, anteriormente) você encontraria os produtos da meiose?
9. Se você presumir que a Figura 2.15 também se aplica a camundongos e você irradiar os espermatozoides do macho com raios X (que sabidamente inativam os genes), qual fenótipo você procuraria na progênie com a finalidade de encontrar casos de camundongos com um gene *SRY* inativado?
10. Na Figura 2.17, como a proporção de 3:1 na grade na parte inferior à esquerda difere das proporções de 3:1 obtidas por Mendel?
11. Na Figura 2.19, presuma que o heredograma seja em relação a camundongos, nos quais qualquer cruzamento escolhido pode ser realizado. Se você cruzar IV-1 com IV-3, qual é a probabilidade de que o primeiro filhote apresente o fenótipo recessivo?
12. Na sua opinião, qual parte do heredograma na Figura 2.23 demonstra melhor a primeira lei de Mendel?
13. O heredograma na Figura 2.31 poderia ser explicado como um distúrbio

autossômico dominante? Explique.

PROBLEMAS BÁSICOS

14. Crie uma sentença que inclua as palavras *cromossomo*, *genes* e *genoma*.

15. Ervilhas (*Pisum sativum*) são diploides e $2n = 14$. Em *Neurospora*, o fungo haploide, $n = 7$. Se você precisasse isolar o DNA genômico de ambas as espécies e utilizasse eletroforese para separar as moléculas do DNA de acordo com o tamanho, quantas bandas de DNA distintas seriam visíveis em cada espécie?

16. A fava (*Vicia faba*) é diploide e $2n = 18$. Cada conjunto haploide de cromossomos contém aproximadamente 4 m de DNA. O tamanho médio de cada cromossomo durante a metáfase da mitose é de 13 µm. Qual é a proporção de empacotamento média do DNA na metáfase? (Proporção de compressão = Comprimento do cromossomo/comprimento da molécula de DNA). Como essa compressão é realizada?

17. Se denominarmos a quantidade de DNA por genoma "*x*", cite uma situação ou situações em organismos diploides na(s) qual(is) a quantidade de DNA por célula seja:

a. *x*.

b. $2x$.

c. $4x$.

18. Cite a função-chave da mitose.

19. Cite as duas funções-chave da meiose.

20. Desenhe um sistema de divisão nuclear diferente que alcance o mesmo desfecho da meiose.

21. Em uma possível situação futura, a fertilidade masculina cai para zero, mas, felizmente, os cientistas desenvolvem um modo para que as mulheres engravidem. Os meiócitos são convertidos diretamente (sem ser submetidos à meiose) em zigotos, que se implantam do modo habitual. Quais seriam os efeitos a curto e longo prazos em uma referida sociedade?

22. De quais modos a segunda divisão da meiose difere da mitose?

23. Crie mnemônicos para relembrar os cinco estágios da prófase I da meiose e os quatro estágios da mitose.

24. Em uma tentativa de simplificar a meiose para o benefício dos estudantes, cientistas malucos desenvolvem um modo de evitar a fase S pré-meiótica e realizá-la com apenas uma divisão, incluindo pareamento, *crossing over* e segregação. Esse sistema funcionaria e os produtos de um referido sistema difeririam daqueles do sistema atual?

25. Theodor Boveri disse: "O núcleo não se divide; ele é dividido." O que ele estava mencionando?

26. Francis Galton, um geneticista da era pré-mendeliana, elaborou o princípio de que metade de nossa constituição genética é derivada de cada um dos pais, um quarto de cada um dos avós, um oitavo de cada um dos bisavós e assim por diante. Ele estava certo? Explique.

27. Se os filhos obtêm metade de seus genes de um dos genitores e metade do outro, porque eles não são irmãos idênticos?

28. Declare onde as células se dividem mitoticamente e onde elas se dividem meioticamente em uma samambaia, um musgo, uma planta florescente, um pinheiro, um cogumelo, uma rã, uma borboleta e um caracol.

29. As células humanas normalmente têm 46 cromossomos. Em relação a cada um dos estágios a seguir, declare o número de moléculas de DNA nuclear presentes em uma célula humana:

a. Metáfase da mitose.

b. Metáfase I da meiose.

c. Telófase da mitose.

d. Telófase I da meiose.

e. Telófase II da meiose.

30. Quatro dos eventos a seguir fazem parte tanto da meiose quanto da mitose, mas apenas um é meiótico. Qual?

a. Formação das cromátides.

b. Formação do fuso.

c. Condensação cromossômica.

d. Movimentação dos cromossomos para os polos.

e. Sinapse.

31. No milho, o alelo *f′* causa um endosperma farinhento e o alelo *f″* causa um endosperma rígido. No cruzamento ♀ *f′*/*f′* × ♂ *f″*/*f″*, todos os endospermas da progênie são farinhentos, mas no cruzamento recíproco, todos os endospermas da progênie são rígidos. Qual é uma possível explicação? (Verifique a legenda da Figura 2.7.)

32. Qual é a primeira lei de Mendel?

33. Se você tivesse uma mosca-das-frutas (*Drosophila melanogaster*) que fosse do fenótipo *A*, qual cruzamento você realizaria para determinar se o genótipo da mosca era *A*/*A* ou *A*/*a*?

34. Ao examinar uma grande amostra de colônias de leveduras em uma placa de Petri, um geneticista observa uma colônia de aspecto anormal, que é muito pequena. Essa pequena colônia foi cruzada com o tipo selvagem e os produtos da meiose (ascósporos) foram espalhados sobre uma placa para produzir colônias. No total, houve 188 colônias do tipo selvagem (tamanho normal) e 180 colônias pequenas.

a. O que pode ser deduzido a partir desses resultados a respeito da herança do fenótipo da colônia pequena? (Invente símbolos genéticos.)

b. Qual seria o aspecto de um asco desse cruzamento?

35. Duas cobaias pretas foram cruzadas e ao longo de diversos anos produziram 29 descendentes pretos e 9 brancos. Explique esses resultados, fornecendo os genótipos dos genitores e da progênie.

36. Em um fungo com quatro ascósporos, um alelo mutante *lys*-5 faz com que os ascósporos que contêm o alelo sejam brancos, enquanto o alelo do tipo selvagem *lys*-5^+ resulta em ascósporos pretos. (Ascósporos são os esporos que constituem os quatro produtos da meiose.) Desenhe um asco de cada um dos cruzamentos a seguir:

a. *lys*-5 × *lys*-5^+.

b. *lys*-5 × *lys*-5.

c. *lys*-5^+ × *lys*-5^+.

37. Em relação a um determinado gene em um organismo diploide, 8 unidades de produto proteico são necessárias para a função normal. Cada alelo do tipo selvagem produz cinco unidades.

a. Se uma mutação criar um alelo nulo, você acredita que esse alelo será recessivo ou dominante?

b. Quais presunções precisam ser feitas para responder a parte *a*?

38. Uma colônia de *Neurospora* sobre uma placa aparentava ser esparsa (densidade baixa) em comparação às outras colônias na placa. Acreditou-se que essa colônia fosse uma possível mutante e então ela foi removida e cruzada com um tipo selvagem do tipo reprodutivo oposto. A partir desse cruzamento, foi obtida uma progênie de 100 ascósporos. Nenhuma das colônias desses ascósporos era esparsa, todas aparentando serem normais. Qual é a explicação mais simples desse resultado? Como você testaria a sua explicação? (Nota: *Neurospora* é haploide.)

39. A partir de uma triagem em grande escala de muitas *Collinsia grandiflora*, foi descoberta uma planta com três cotilédones (normalmente existem dois cotilédones). Essa planta foi cruzada com uma planta do tipo selvagem pura normal e foram plantadas 600 sementes desse cruzamento. Houve 298 plantas com 2 cotilédones e 302 com 3 cotilédones. O que pode ser deduzido a respeito da herança dos três cotilédones? Invente símbolos gênicos como parte da sua explicação.

40. Na planta *Arabidopsis thaliana*, um geneticista está interessado no desenvolvimento de tricomas (pequenas projeções). Uma grande triagem resulta em duas plantas mutantes (A e B) que não apresentam tricomas e esses mutantes aparentam ser possivelmente úteis no estudo do desenvolvimento de tricomas. (Se eles fossem determinados por mutações monogênicas, o achado das funções normal e anormal desses genes seria instrutivo.) Cada planta é cruzada com o tipo selvagem; em ambos os casos,

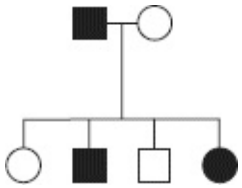
a próxima geração (F_1) apresentou tricomas normais. Quando as plantas da F_1 foram autofecundadas, a F_2 resultante foi como segue:

F_2 do mutante A: 602 normais; 198 sem tricomas.
F_2 do mutante B: 267 normais; 93 sem tricomas.

a. O que esses resultados demonstram? Inclua os genótipos propostos de todas as plantas em sua resposta.
b. De acordo com sua explicação para a parte *a*, é possível prever com confiança a F_1 do cruzamento do mutante A original com o mutante B original?

41. Você tem três dados: um vermelho (R), um verde (G) e um azul (B). Quando todos os três dados são jogados ao mesmo tempo, calcule a probabilidade dos desfechos a seguir:
a. 6 (R), 6 (G), 6 (B)
b. 6 (R), 5 (G), 6 (B)
c. 6 (R), 5 (G), 4 (B)
d. Absolutamente nenhum seis
e. Um número diferente em cada dado.

42. No heredograma a seguir, os símbolos pretos representam indivíduos com uma doença sanguínea muito rara.



Se você não tivesse nenhuma outra informação para prosseguir, você acharia mais provável que a doença fosse dominante ou recessiva? Dê os seus motivos.

43. **a.** A capacidade de degustar a substância química feniltiocarbamida é um fenótipo autossômico dominante e a incapacidade de degustá-la é

recessiva. Se uma mulher degustadora com um pai não degustador se casar com um homem degustador que em um casamento anterior teve uma filha não degustadora, qual é a probabilidade de que o seu primeiro filho seja:

(1) Uma menina não degustadora.
(2) Uma menina degustadora.
(3) Um menino degustador.

b. Qual é a probabilidade de que os seus dois primeiros filhos de qualquer sexo sejam degustadores?

Como solucionar o problema 44

John e Martha estão pensando em ter filhos, mas o irmão de John apresenta galactosemia (uma doença recessiva autossômica) e a bisavó de Martha também apresentava galactosemia. Martha tem uma irmã que tem três filhos, nenhum dos quais apresenta galactosemia. Qual é a probabilidade de que o primeiro filho de John e Martha apresente galactosemia?

1. O problema pode ser reformulado como um heredograma? Caso afirmativo, escreva um.
2. Partes do problema podem ser reformuladas por meio da utilização de quadrados de Punnett?
3. Partes do problema podem ser reformuladas por meio da utilização de diagramas ramificados?
4. No heredograma, identifique um cruzamento que ilustre a primeira lei de Mendel.
5. Defina todos os termos científicos no problema e procure quaisquer outros termos a respeito dos quais você não tenha certeza.
6. Quais presunções precisam ser feitas ao responder esse problema?
7. Quais familiares não mencionados precisam ser considerados? Por quê?
8. Quais regras estatísticas poderiam ser relevantes e em quais situações elas podem ser aplicadas? As referidas situações existem neste problema?
9. Quais são duas generalidades a respeito das doenças autossômicas

recessivas em populações humanas?

10. Qual é a relevância da raridade do fenótipo em estudo na análise de heredogramas em geral e o que pode ser inferido neste problema?
11. Nessa família, os genótipos de quais pessoas são certos e de quais pessoas são incertos?
12. De que modo o lado de John do heredograma é diferente daquele do lado de Martha? Como essa diferença afeta os seus cálculos?
13. Existem quaisquer informações irrelevantes no problema conforme formulado?
14. De que modo a solução desse tipo de problema é similar à de problemas que você já resolveu com sucesso? De que modo ela é diferente?
15. Você pode criar uma breve história com base no dilema humano neste problema?

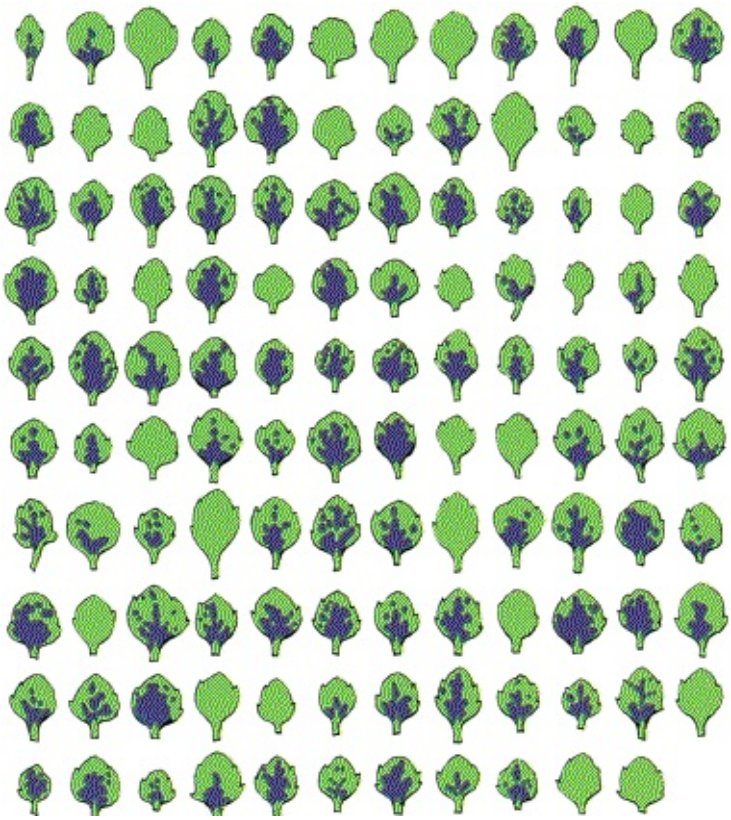
Agora tente solucionar o problema. Se você não conseguir, tente identificar o obstáculo e escreva uma sentença ou duas que descrevam a sua dificuldade. Em seguida, retorne às questões de expansão e verifique se qualquer uma delas se relaciona à sua dificuldade.

45. O gado Holstein normalmente é preto e branco. Um touro branco e preto esplêndido, Charlie, foi adquirido por um fazendeiro por US$ 100.000,00. Toda a progênie gerada por Charlie foi de aspecto normal. Entretanto, determinados pares de sua progênie, quando entrecruzados, produziram progênie vermelha e branca numa frequência de aproximadamente 25%. Charlie logo foi removido das listas de reprodutores dos criadores de Holstein. Utilize símbolos para explicar precisamente o motivo.

46. Suponha que o marido e a esposa sejam, ambos, heterozigotos em relação a um alelo recessivo para o albinismo. Se eles tiverem gêmeos dizigóticos (dois ovócitos), qual é a probabilidade de que ambos os gêmeos tenham o mesmo fenótipo em relação à pigmentação?

47. A planta *Omphalodes verna* cresce em locais do Canadá como a ilha de Vancouver e a parte continental mais baixa da Colúmbia Britânica. As

populações são dimórficas para manchas roxas nas folhas — algumas plantas apresentam manchas e outras não. Próximo de Nanaimo, uma planta na natureza apresentava folhas manchadas. Essa planta, que ainda não havia florescido, foi colhida e levada para um laboratório, onde se possibilitou sua autopolinização. As sementes foram coletadas e cultivadas, dando origem à progênie. Uma folha aleatoriamente selecionada (mas típica) de cada uma da progênie está demonstrada na ilustração a seguir:



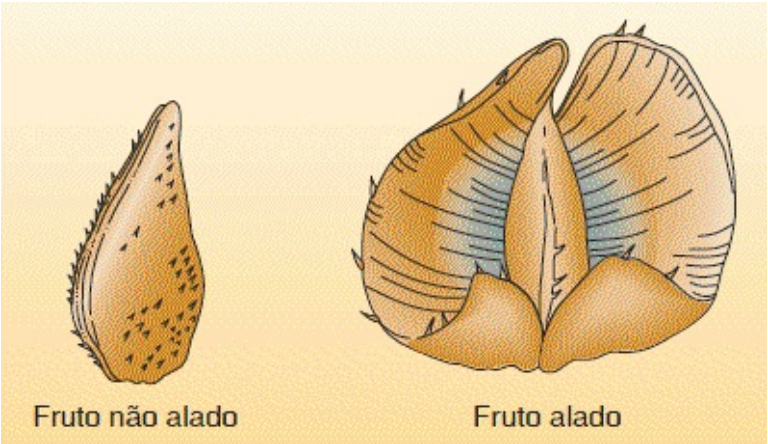
a. Formule uma hipótese genética concisa para explicar esses resultados. Explique todos os símbolos e demonstre todas as classes genotípicas (e o genótipo da planta original).

b. Como você testaria a sua hipótese? Seja específico.

48. Pode chegar a ser comprovado que um animal não é um portador de um alelo recessivo (ou seja, não é um heterozigoto em relação a um determinado gene)? Explique.

49. Na natureza, a planta *Plectritis congesta* é dimórfica para o formato do

fruto; ou seja, plantas individuais contêm frutos não alados ou alados, conforme demonstrado na ilustração.



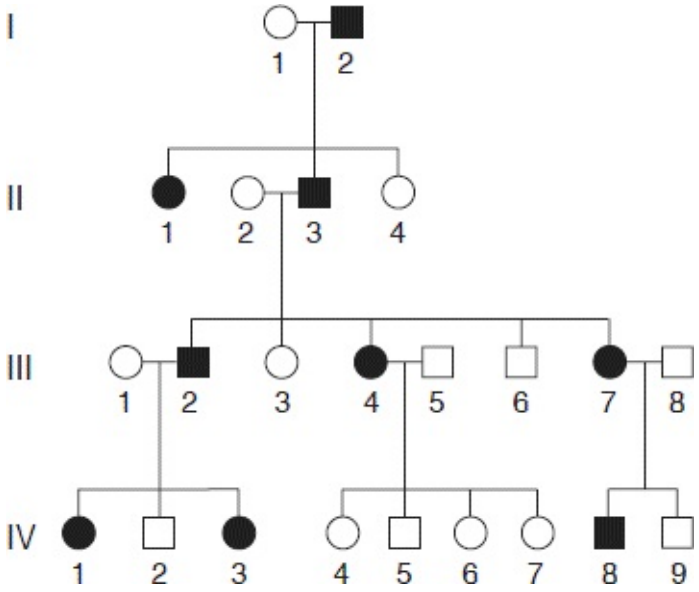
Foram coletadas plantas da natureza antes da floração, as quais foram cruzadas ou autofecundadas, com os resultados a seguir:

	Número de progênie	
Polinização	Alados	Não alados
Alados (autofecundadas)	91	1*
Alados (autofecundadas)	90	30
Não alados (autofecundadas)	4*	80
Alados × Não alados	161	0
Alados × Não alados	29	31
Alados × Não alados	46	0
Alados × Alados	44	0

*O fenótipo provavelmente tem uma explicação não genética.

Interprete esses resultados e derive o modo de herança desses fenótipos do formato do fruto. Utilize símbolos. Qual você acredita que seja a explicação não genética para os fenótipos marcados pelos asteriscos na tabela?

50. O heredograma a seguir é em relação a um distúrbio de pele hereditário raro, porém relativamente brando.

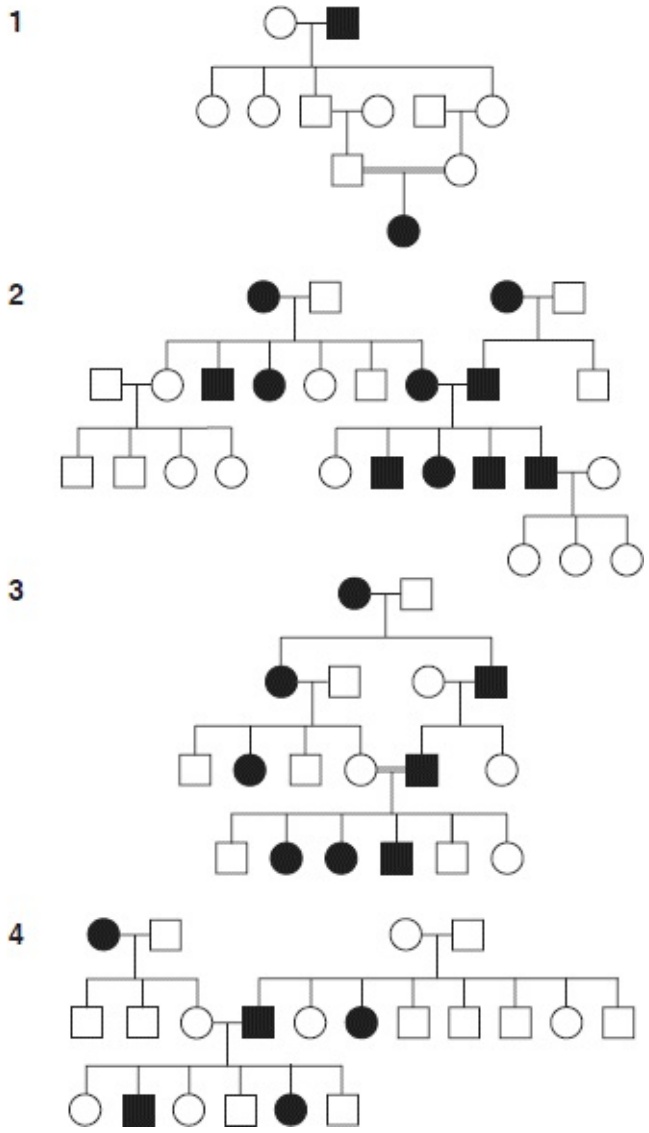


a. Como o distúrbio é herdado? Declare os motivos da sua resposta.

b. Forneça os genótipos para tantos indivíduos no heredograma quanto possível. (Invente os seus próprios símbolos alélicos definidos.)

c. Considere os quatro filhos não afetados dos genitores III-4 e III-5. Em todas as progênies com quatro filhos de genitores desses genótipos, qual proporção se espera que contenha todos os filhos não afetados?

51. Quatro heredogramas humanos são demonstrados na ilustração a seguir. Os símbolos pretos representam um fenótipo anormal herdado de modo mendeliano simples.



a. Em relação a cada heredograma, declare se a condição anormal é dominante ou recessiva. Tente declarar a lógica por trás da sua resposta.
b. Em relação a cada heredograma, descreva os genótipos de tantas pessoas quanto possível.

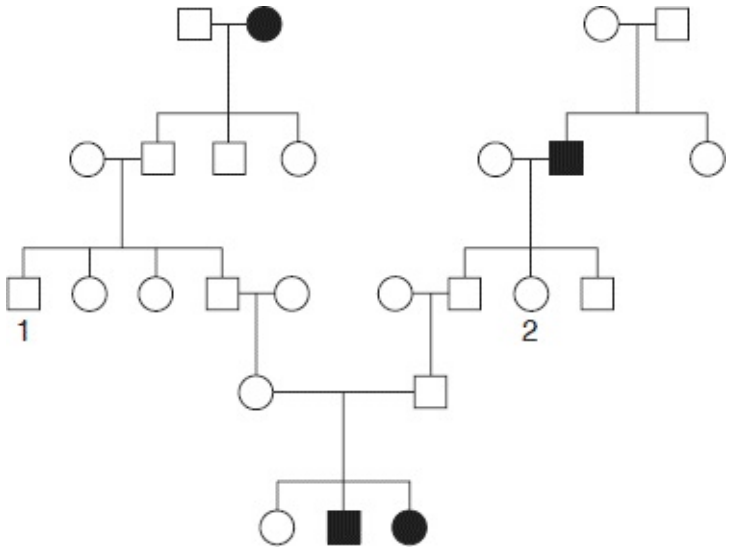
52. A doença de Tay-Sachs é uma doença humana rara, na qual substâncias tóxicas acumulam-se nas células nervosas. O alelo recessivo responsável pela doença é herdado de modo mendeliano simples. Por motivos desconhecidos, o alelo é mais comum em populações de judeus asquenaze do Leste Europeu. Uma mulher está planejando se casar com seu primo em primeiro grau, mas o casal descobre que a irmã de seu avô em comum

morreu no primeiro ano de vida de doença de Tay-Sachs.

a. Desenhe as partes relevantes do heredograma e demonstre todos os genótipos tão completamente quanto possível.

b. Qual é a probabilidade de que o primeiro filho dos primos apresente doença de Tay-Sachs, presumindo que todas as pessoas que se casam dentro da família sejam homozigotas normais?

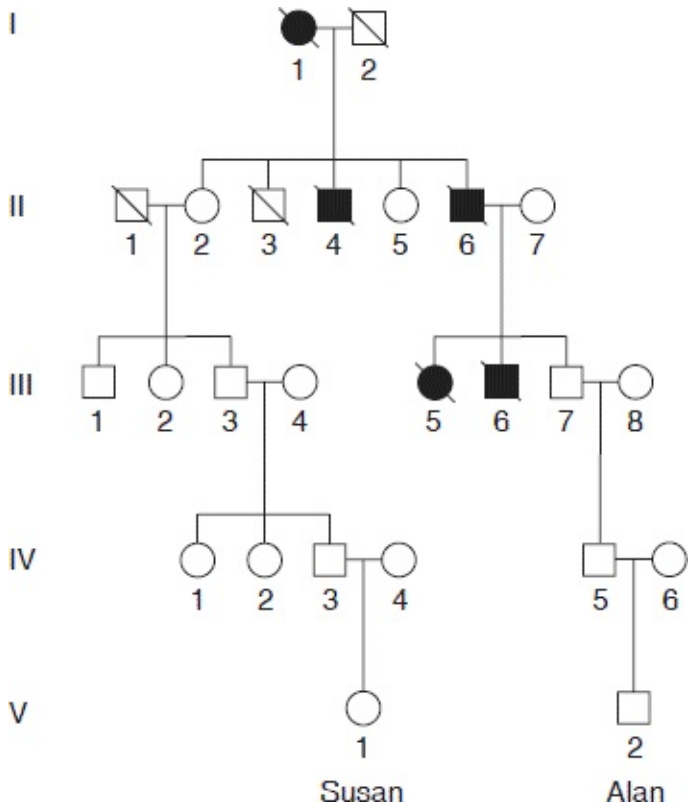
53. O heredograma a seguir foi obtido em relação a uma doença renal rara.



a. Deduza a herança dessa condição, declarando os seus motivos.

b. Se as pessoas 1 e 2 se casarem, qual é a probabilidade de que o seu primeiro filho apresente a doença renal?

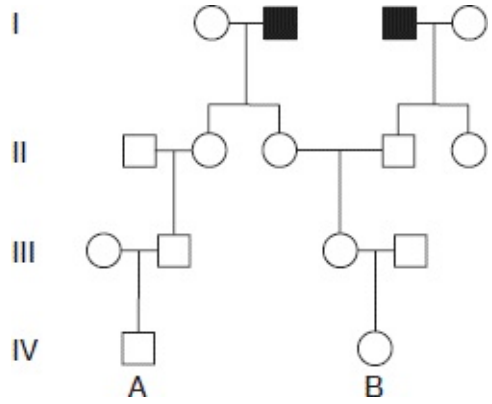
54. O heredograma a seguir é em relação à doença de Huntington, um distúrbio do sistema nervoso de início tardio. As barras indicam familiares falecidos.



a. Esse heredograma é compatível com o modo de herança em relação à doença de Huntington mencionado no capítulo?

b. Considere duas crianças recém-nascidas nos dois braços do heredograma, Susan no braço esquerdo e Alan no braço direito. Estude o gráfico na Figura 2.24 e forme uma opinião sobre a probabilidade de que eles desenvolvam doença de Huntington. Presuma, para fins de discussão, que os genitores tiveram filhos aos 25 anos de idade.

55. Considere o heredograma que acompanha de uma doença autossômica recessiva rara, a PKU.



a. Liste os genótipos de tantos dos familiares quanto possível.

b. Se as pessoas A e B se casarem, qual é a probabilidade de que seu primeiro filho apresente PKU?

c. Se o seu primeiro filho for normal, qual é a probabilidade de que o seu segundo filho apresente PKU?

d. Se o seu primeiro filho apresentar a doença, qual é a probabilidade de que o seu segundo filho não seja afetado?

(Presuma que todas as pessoas que se casam no heredograma não possuam o alelo anormal.)

56. Um homem apresenta lóbulos da orelha presos, enquanto sua esposa apresenta lóbulos da orelha livres. Seu primeiro filho, um menino, apresenta lóbulos da orelha presos.

a. Se for presumido que a diferença fenotípica ocorre em virtude de dois alelos de um gene único, é possível que o gene esteja ligado ao X?

b. É possível decidir se os lóbulos da orelha presos são uma característica dominante ou recessiva?

57. Um alelo recessivo raro herdado de modo mendeliano causa a doença fibrose cística. Um homem fenotipicamente normal cujo pai apresentava fibrose cística se casa com uma mulher fenotipicamente normal de fora da família e o casal considera ter um filho.

a. Desenhe o heredograma, como descrito.

b. Se a frequência na população de heterozigotos em relação à fibrose cística é de 1 em 50, qual é a chance de que o primeiro filho do casal

apresente fibrose cística?

c. Se o primeiro filho apresentar fibrose cística, qual é a probabilidade de que o segundo filho seja normal?

58. O alelo *c* causa albinismo em camundongos (*C* faz com que os camundongos sejam pretos). O cruzamento $C/C \times c/c$ produz uma progênie de 10. Qual é a probabilidade de que toda ela seja preta?

59. O alelo recessivo *s* faz com que a *Drosophila* apresentem asas pequenas e o alelo s^+ causa asas normais. Esse gene sabidamente é ligado ao X. Se um macho com asas pequenas for cruzado com uma fêmea homozigota do tipo selvagem, qual proporção de moscas normais e com asas pequenas pode ser esperada em cada sexo na F_1? Se as moscas da F_1 forem entrecruzadas, quais proporções da progênie F_2 são esperadas? Quais proporções de progênie são previstas se as fêmeas da F_1 forem retrocruzadas com seu pai?

60. Um alelo dominante ligado ao X causa hipofosfatemia em humanos. Um homem com hipofosfatemia se casa com uma mulher normal. Qual proporção de seus filhos apresentará hipofosfatemia?

61. A distrofia muscular de Duchenne é ligada ao sexo e normalmente afeta apenas homens. As vítimas da doença se tornam progressivamente mais fracas, e a doença se manifesta precocemente.

a. Qual é a probabilidade de que uma mulher cujo irmão apresenta a distrofia de Duchenne tenha um filho afetado?

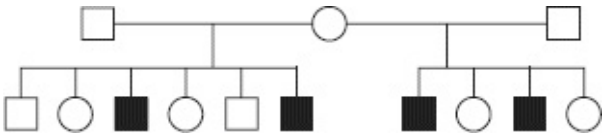
b. Se o irmão da sua mãe (seu tio) teve distrofia de Duchenne, qual é a probabilidade de que você tenha recebido o alelo?

c. Se o irmão do seu pai teve a doença, qual é a probabilidade de que você tenha recebido o alelo?

62. Um homem e uma mulher casados recentemente descobrem que cada um teve um tio com alcaptonúria, uma doença rara causada por um alelo autossômico recessivo de um gene único. Eles estão para ter o seu primeiro bebê. Qual é a probabilidade de que seu filho apresente alcaptonúria?

63. O heredograma a seguir se refere a uma anormalidade dentária hereditária

rara, a amelogênese imperfeita.



a. Qual modo de herança explica melhor a transmissão desse traço?

b. Escreva os genótipos de todos os familiares de acordo com a sua hipótese.

64. Um casal que está para se casar descobre, ao estudar seus históricos familiares, que, em *ambas* as suas famílias, seus avós não afetados tinham irmãos com fibrose cística (uma doença autossômica recessiva rara).

a. Se o casal se casar e tiver um filho, qual é a probabilidade de que o filho apresente fibrose cística?

b. Se eles tiverem quatro filhos, qual é a chance de que os filhos apresentem a proporção mendeliana precisa de 3:1 de normais:fibrose cística?

c. Se o primeiro filho deles apresentar fibrose cística, qual é a probabilidade de que os próximos três filhos sejam normais?

65. Um alelo *c* recessivo ligado ao sexo produz daltonismo verde-vermelho em humanos. Uma mulher normal cujo pai era daltônico se casa com um homem daltônico.

a. Quais genótipos são possíveis em relação à mãe do homem daltônico?

b. Quais são as chances de que o primeiro filho desse casamento seja um menino daltônico?

c. Das meninas geradas por esses genitores, pode-se esperar que qual proporção seja daltônica?

d. De todos os filhos (sexo não especificado) desses genitores, pode-se esperar que qual proporção apresente visão normal para cores?

66. Gatos domésticos machos são pretos ou laranja; as fêmeas são pretas, laranja ou cálicas (tricolores).

a. Se esses fenótipos de cor de pelagem forem controlados por um gene

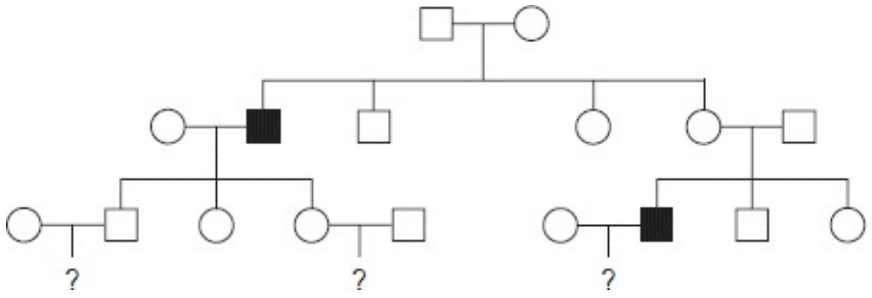
ligado ao sexo, como essas observações podem ser explicadas?

b. Com a utilização de símbolos apropriados, determine os fenótipos esperados na progênie de um cruzamento entre uma fêmea laranja e um macho preto.

c. Metade das fêmeas produzidas por um determinado tipo de cruzamento é cálica e metade é preta; metade dos machos é laranja e metade é preta. De que cores são os machos e as fêmeas parentais nesse tipo de cruzamento?

d. Outro tipo de cruzamento produz progênie nas proporções a seguir: um quarto de machos laranja, um quarto de fêmeas laranja, um quarto de machos pretos e um quarto de fêmeas cálicas. De que cores são os machos e as fêmeas parentais nesse tipo de cruzamento?

67. O heredograma a seguir refere-se a uma determinada doença rara que é incapacitante, porém não fatal.



a. Determine o modo de herança mais provável dessa doença.

b. Escreva o genótipo de cada familiar de acordo com o seu modo de herança proposto.

c. Se você fosse o médico dessa família, como informaria os três casais na terceira geração a respeito da probabilidade de terem um filho afetado?

68. No milho, o alelo *s* causa endosperma açucarado, enquanto *S* causa endosperma amiláceo. Quais genótipos de endosperma resultam de cada um dos cruzamentos a seguir?

a. Sexo feminino *s*/*s* × sexo masculino *S*/*S*.

b. Sexo feminino *S*/*S* × sexo masculino *s*/*s*.

c. Sexo feminino *S*/*s* × sexo masculino *S*/*s*.

69. Uma geneticista de plantas possui duas linhagens puras, uma com pétalas roxas e uma com pétalas azuis. Ela formula a hipótese de que a diferença fenotípica ocorre em virtude de dois alelos de um gene. Para testar essa ideia, ela pretende observar uma proporção de 3:1 na F_2. Ela cruza as linhagens e obtém toda a progênie F_1 roxa. As plantas da F_1 são autopolinizadas e são obtidas 400 plantas da F_2. Dessas plantas da F_2, 320 são roxas e 80 são azuis. Esses resultados correspondem bem à sua hipótese? Caso negativo, sugira o motivo.

70. O avô de um homem apresenta galactosemia, uma doença autossômica recessiva rara causada pela incapacidade de processar a galactose, que leva à disfunção muscular, nervosa e renal. O homem se casou com uma mulher cuja irmã apresentava galactosemia. A mulher agora está grávida de seu primeiro filho.

a. Desenhe o *heredograma*, conforme descrito.

b. Qual é a probabilidade de que esse filho apresente galactosemia?

c. Se o primeiro filho apresentar galactosemia, qual é a probabilidade de que um segundo filho a apresente?

PROBLEMAS DESAFIADORES

71. Um geneticista que trabalha com ervilhas tem uma única planta mono-híbrida *Y/y* (amarela) e, a partir da autofecundação dessa planta, deseja produzir uma planta de genótipo *y/y* para ser utilizada como testador. Quantas plantas da progênie precisam ser cultivadas para se ter 95% de certeza de que se obterá no mínimo uma na amostra?

72. Um polimorfismo curioso em populações humanas está relacionado com a capacidade de dobrar as laterais da língua para formar uma depressão (“rolamento da língua”). Algumas pessoas conseguem realizar esse truque, enquanto outras simplesmente não conseguem. Portanto, esse é um exemplo de um dimorfismo. A sua significância é um completo mistério. Em uma família, um menino era incapaz de dobrar a sua língua, mas, para sua grande decepção, sua irmã conseguia. Além disso, ambos os genitores

conseguiam realizar o rolamento e assim também ambos os avós, um tio paterno e uma tia paterna. Uma tia paterna, um tio paterno e um tio materno não conseguiam dobrar suas línguas.

a. Desenhe o heredograma em relação a essa família, definindo os seus símbolos claramente e deduza os genótipos de tantos membros quanto possível.

b. O heredograma que você desenhou é típico da herança de dobrar a língua e levou os geneticistas a avaliarem o mecanismo de herança que sem dúvida você deduziu. Entretanto, em um estudo de 33 pares de gêmeos idênticos, ambos os membros de 18 pares conseguiam fazê-lo, nenhum dos membros de 8 pares conseguia e um dos gêmeos em 7 pares conseguia realizar o rolamento, mas o outro não. Tendo em vista que gêmeos idênticos são derivados da divisão de um ovócito fertilizado em dois embriões, os membros de um par têm de ser geneticamente idênticos. Como a existência dos sete pares discordantes pode ser conciliada com a sua explicação genética do heredograma?

73. Os cabelos ruivos ocorrem em famílias, conforme o heredograma a seguir demonstra. (Dados de heredograma de W. R. Singleton e B. Ellis, *Journal of Heredity* 55, 1964, 261.)

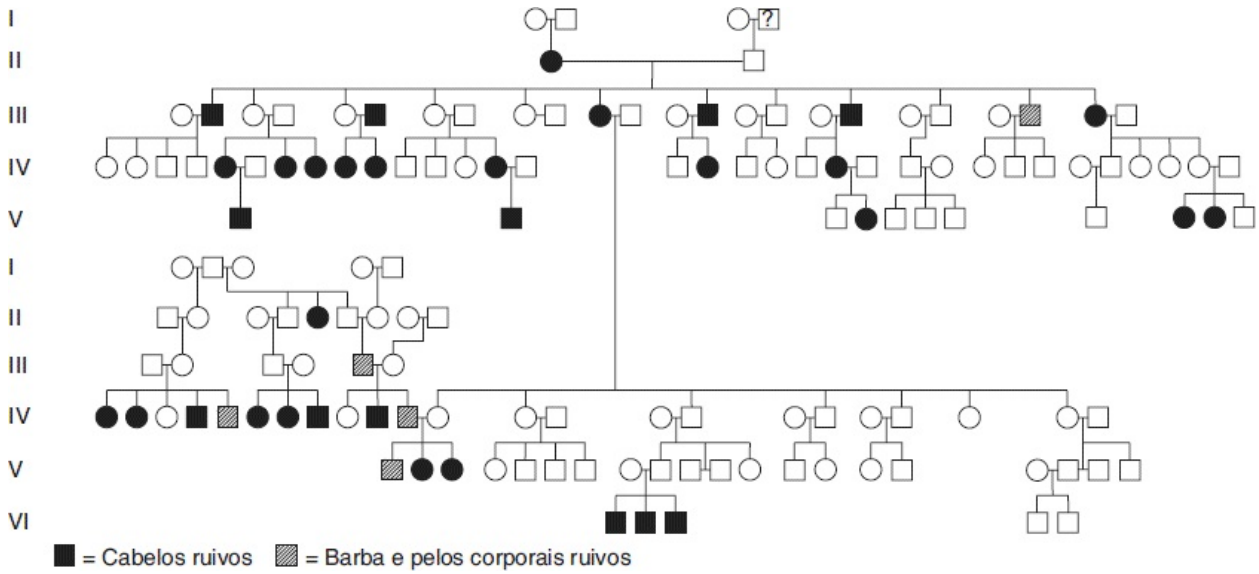
a. O padrão de herança nesse heredograma sugere que os cabelos ruivos poderiam ser causados por um alelo dominante ou recessivo de um gene que é herdado de modo mendeliano simples?

b. Você acredita que o alelo de cabelos ruivos seja comum ou raro na população como um todo?

74. Quando muitas famílias foram testadas em relação à capacidade de degustar a substância química feniltiocarbamida, os cruzamentos foram agrupados em três tipos e a progênie foi totalizada, com os resultados demonstrados a seguir:

Filhos

Genitores	Número de famílias	Degustadores	Não degustadores
Degustador × Degustador	425	929	130
Degustador × Não degustador	289	483	278
Não degustador × Não degustador	86	5	218



Com a presunção de que a degustação da PTC seja dominante (*P*) e a não degustação seja recessiva (*p*), como as proporções da progênie em cada um dos três tipos de cruzamentos podem ser explicadas?

75. Uma condição conhecida como ictiose *hystrix gravior* apareceu em um menino no início do século 18. Sua pele se tornou muito espessa e formou espinhos, que se desprendiam em intervalos. Quando cresceu, esse “homem

porco-espinho” se casou e teve seis filhos, todos os quais apresentavam essa condição, e diversas filhas, todas normais. Durante quatro gerações, essa condição foi transmitida de pai para filho. A partir dessa evidência, o que você pode postular a respeito da localização do gene?

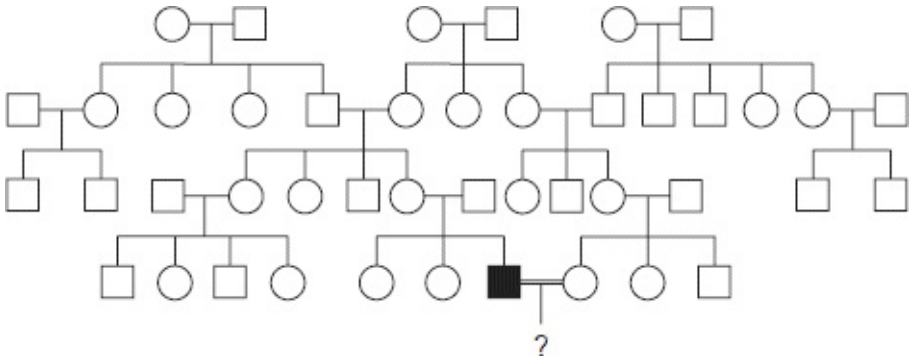
76. A mariposa *Abraxas* do tipo selvagem (W) apresenta grandes manchas em suas asas, mas a forma lacticolor (L) dessa espécie apresenta manchas muito pequenas. Foram realizados cruzamentos entre linhagens que diferem nessa característica, com os resultados a seguir:

	Genitores		*Progênie*	
Cruzamento	♀	♂	F_1	F_2
1	L	W	♀ W	♀ $\frac{1}{2}$ L, $\frac{1}{2}$ W
			♂ W	♂ W
2	W	L	♀ L	♀ $\frac{1}{2}$ W, $\frac{1}{2}$ L
			♂ W	♂ $\frac{1}{2}$ W, $\frac{1}{2}$ L

Forneça uma explicação genética clara dos resultados nesses dois cruzamentos, demonstrando os genótipos de todas as mariposas.

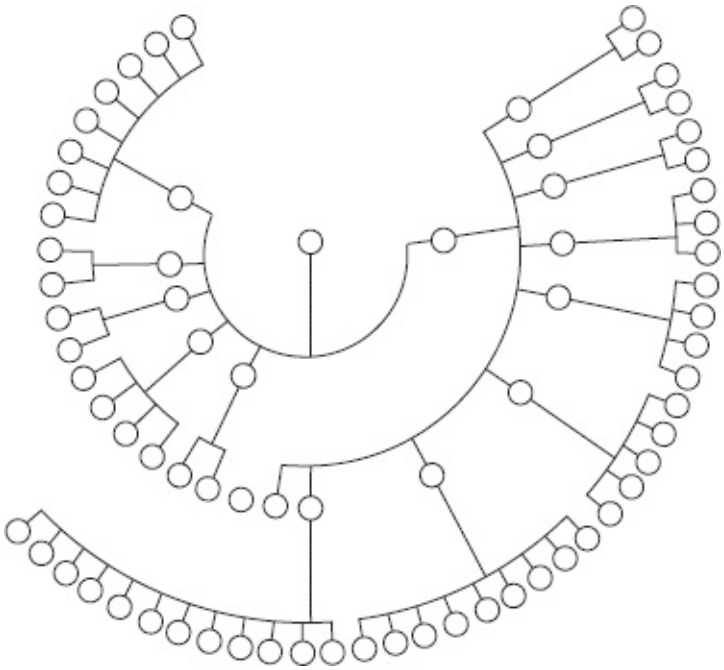
77. O heredograma na parte inferior da página demonstra a herança de uma doença humana rara. O padrão é mais bem-explicado como causado por um alelo recessivo ligado ao X ou por um alelo autossômico dominante com expressão limitada aos homens? (Dados de heredograma de J. F. Crow, *Genetics Notes,* 6th ed. Copyright 1967 por Burgess Publishing Co., Minneapolis.)

78. Um determinado tipo de surdez em humanos é herdado como um traço recessivo ligado ao X. Um homem com esse tipo de surdez se casou com uma mulher normal, e eles estão esperando um filho. Eles descobrem que são parentes afastados. Parte do heredograma está demonstrada aqui.



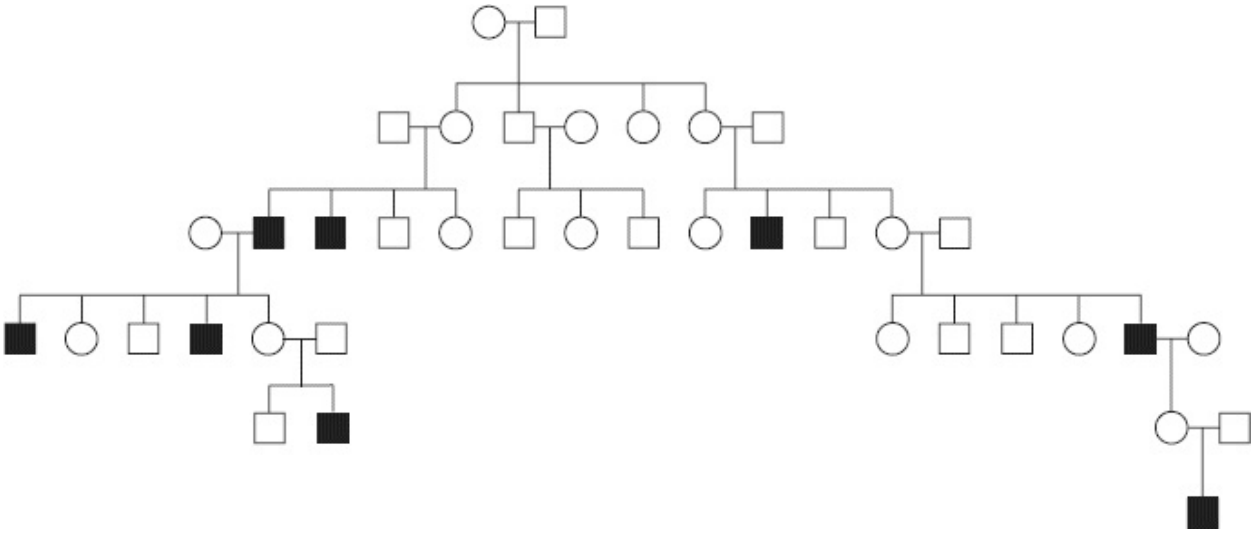
Como você alertaria os pais a respeito da probabilidade de seu filho ser um menino surdo, uma menina surda, um menino normal ou uma menina normal? Assegure-se de declarar todas as presunções que você faz.

79. O heredograma a seguir demonstra um padrão de herança muito incomum, que de fato existiu. Toda a progênie é mostrada, mas os pais em cada cruzamento foram omitidos para chamar a atenção para o padrão extraordinário.



a. Declare de modo conciso o que é incomum a respeito desse heredograma.

b. O padrão pode ser explicado por meio da herança mendeliana?

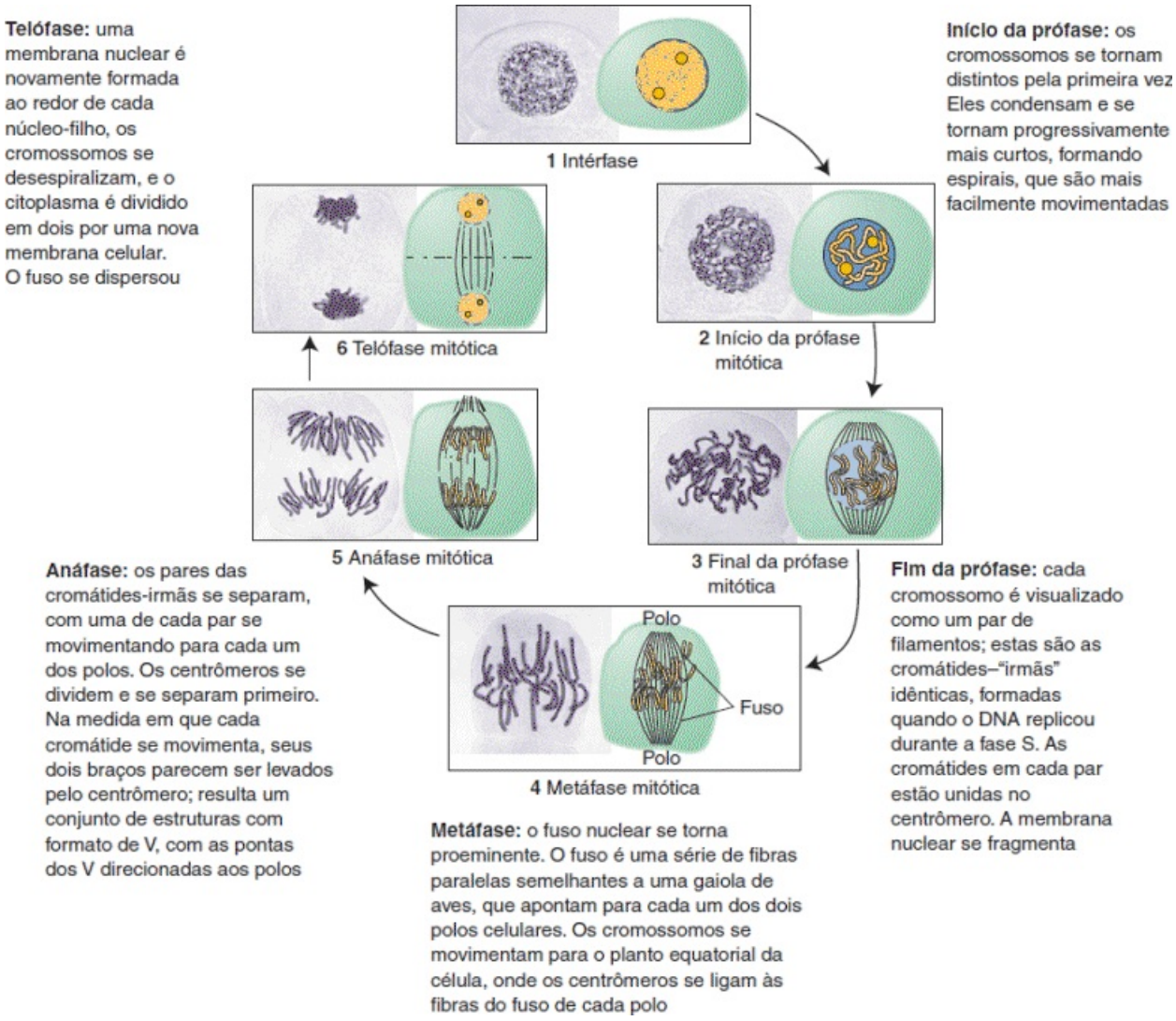


APÊNDICE 2.1 Estágios da mitose

A mitose normalmente ocupa apenas uma pequena proporção do ciclo celular, aproximadamente 5 a 10%. O tempo remanescente é a intérfase, composta pelos estágios G1, S e G2. O DNA é replicado durante a fase S, embora o DNA duplicado não se torne visível até posteriormente na mitose. Os cromossomos não podem ser observados durante a intérfase (ver seguir), principalmente em virtude de estarem em um estado estendido e entrelaçados entre si como um emaranhado de fios.

As fotografias demonstram a mitose nos núcleos de células da extremidade da raiz do *Lilium regale*. (J. McLeish and B. Snoad, *Looking at Chromosomes*. Copyright 1958, St. Martin's, Macmillan.)

As fotografias a seguir demonstram os estágios da mitose nos núcleos de células da extremidade da raiz do lírio real, *Lilium regale*. Em cada um dos estágios, é demonstrada uma fotografia à esquerda e um desenho interpretativo à direita.



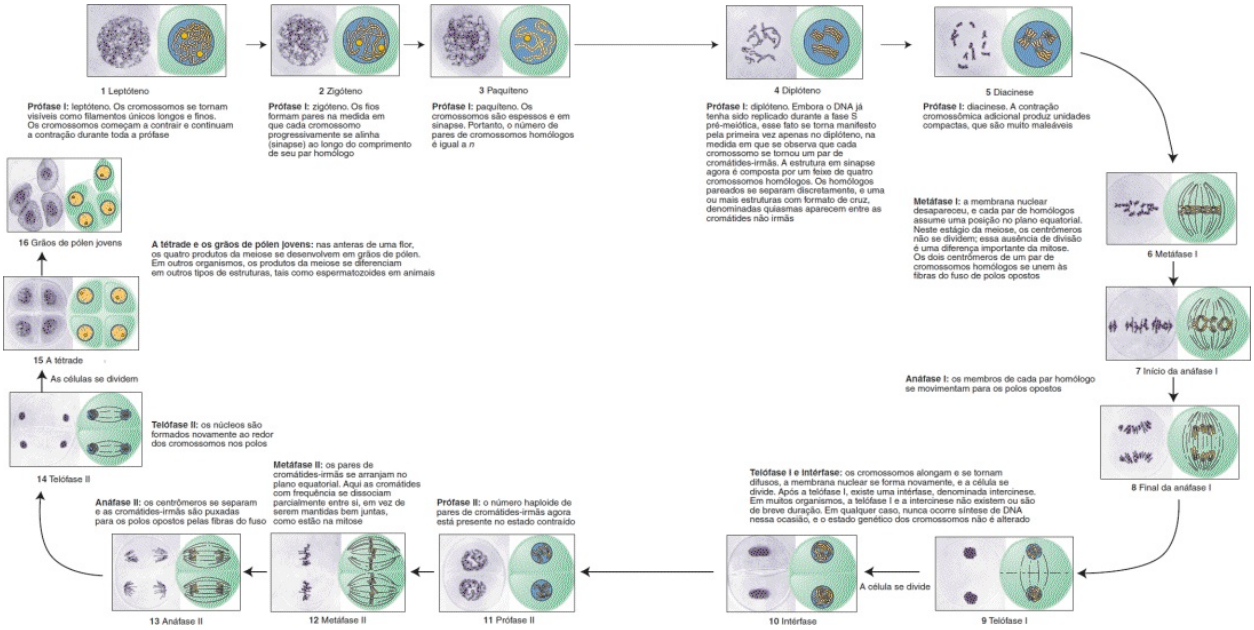
As fotografias demonstram a mitose nos núcleos de células da extremidade da raiz do *Lilium regale*. (J. McLeish and B. Snoad, *Looking at Chromosomes*. Copyright 1958, St. Martin's, Macmillan.)

APÊNDICE 2.2 Estágios da meiose

A meiose é composta por duas divisões nucleares, diferenciadas como meiose I e meiose II, que ocorrem em divisões celulares consecutivas. Cada divisão meiótica é formalmente dividida em prófase, metáfase, anáfase e telófase. Desses estágios, o mais complexo e duradouro é a prófase I, que é dividida em cinco estágios.

As fotografias a seguir demonstram os estágios da meiose nos núcleos de células da extremidade da raiz do lírio real, *Lilium regale*. Em cada um dos

estágios, é demonstrada uma fotografia à esquerda e um desenho interpretativo à direita.



As fotografias demonstram a meiose e a formação do pólen em *Lilium regale*. Nota: para simplificar, são desenhados quiasmas múltiplos entre duas cromátides apenas; na realidade, todas as quatro cromátides podem participar. (J. McLeish and B. Snoad, *Looking at Chromosomes*. Copyright 1958, St. Martin's, Macmillan.)

Segregação Independente de Genes

CAPÍTULO 3

A Revolução Verde na agricultura é promovida pelo plantio disseminado de linhagens superiores de cultivos (tal como o arroz, demonstrado aqui) produzidas por meio da combinação de traços genéticos benéficos. (Jorgen Schytte.)

TÓPICOS

3.1 Lei de Mendel de segregação independente
3.2 A segregação independente
3.3 Base cromossômica da segregação independente
3.4 Herança poligênica

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Em diploides, elaborar experimentos para produzir um di-híbrido e em seguida realizar o seu autocruzamento ou cruzamento-teste
- Em diploides, analisar os fenótipos da progênie de autocruzamentos e cruzamentos-teste de di-híbridos e, a partir desses resultados, avaliar se os dois genes estão se distribuindo de modo independente (o que sugeriria localizações em diferentes cromossomos)
- Em haploides, desenhar experimentos para produzir um di-híbrido diploide temporário *AaBb* e analisar a sua progênie haploide para avaliar se o dois genes estão segregando de modo independente
- Em cruzamentos que envolvem di-híbridos com segregação independente, prever as proporções genotípicas nos produtos meióticos, as proporções genotípicas na progênie e as proporções fenotípicas na progênie
- Utilizar a análise do qui-quadrado para testar se as proporções fenotípicas observadas estão em uma correspondência aceitável com aquelas previstas por segregação independente
- Em diploides, elaborar experimentos para sintetizar linhagens que sejam puras (homozigotas) em relação a dois ou mais genes
- Interpretar as proporções da segregação independente de dois genes em termos do comportamento cromossômico na meiose
- Analisar as proporções da progênie de di-híbridos em termos da frequência de recombinantes (FR) e aplicar a FR diagnóstica para a segregação independente
- Estender os princípios da segregação independente de dois genes para heterozigotos em relação a três ou mais genes
- Estender o princípio da segregação independente para genes múltiplos que contribuem, cada um, para um fenótipo que demonstra distribuição contínua
- Aplicar os critérios diagnósticos para avaliar se as mutações estão em

genes de organelas citoplasmáticas.

Este capítulo é a respeito dos princípios que atuam quando dois ou mais casos de herança monogênica são analisados simultaneamente. Em nenhuma outra área esses princípios foram mais importantes do que cultivo de plantas e na criação de animais na agricultura. Por exemplo, entre os anos de 1960 e 2000, a produção mundial de plantas alimentícias duplicou, marcando uma assim denominada Revolução Verde. O que tornou essa Revolução Verde possível? Em parte, ela ocorreu em virtude da melhora da prática da agricultura, mas mais importante foi o desenvolvimento de genótipos de cultivos superiores por geneticistas de plantas. Esses cultivadores estão em uma busca constante por mutações aleatórias em genes únicos que aumentem significativamente a produção ou o valor nutritivo. Entretanto, as referidas mutações surgem em linhagens diferentes em distintas partes do mundo. Por exemplo, no arroz, um dos principais cultivos alimentícios mundiais, as mutações a seguir foram cruciais na Revolução Verde:

sd1. Esse alelo recessivo resulta em estatura baixa, tornando a planta mais resistente à queda no vento e na chuva; ele também aumenta a quantidade relativa da energia da planta que é dirigida para a semente, a parte que ingerimos.

se1. Esse alelo recessivo altera a necessidade da planta por uma duração específica de período diurno, possibilitando que ela seja cultivada em diferentes latitudes.

Xa4. Esse alelo dominante confere resistência à doença ferrugem bacteriana.

bph2. Esse alelo confere resistência à cigarrinha-castanha (um tipo de inseto).

Snb1. Esse alelo confere tolerância à submersão da planta após chuvas fortes.

Para produzir um genótipo verdadeiramente superior, a combinação dos referidos alelos em uma linhagem é claramente desejável. Para conquistar uma referida combinação, linhagens mutantes têm de ser intercruzadas duas de cada vez. Por

exemplo, um geneticista de plantas poderia iniciar por meio do cruzamento de uma linhagem homozigota para *sd1* com outra homozigota para *Xa4*. A progênie da F_1 desse cruzamento carregaria ambas as mutações, porém em um estado heterozigoto. Entretanto, a maior parte da agricultura utiliza linhagens puras, tendo em vista que elas podem ser propagadas de modo eficiente e distribuídas para os fazendeiros. Para obter uma linhagem duplamente mutante pura *sd1*/*sd1* · *Xa4*/*Xa4*, a F_1 precisaria ser adicionalmente cultivada para possibilitar que os alelos se "distribuíssem" na combinação desejável. Alguns produtos do referido cruzamento estão demonstrados na Figura 3.1. Quais princípios são relevantes aqui? Depende muito se os dois genes estão no mesmo par de cromossomos ou em pares de cromossomos diferentes. No último caso, os pares de cromossomos atuam de modo independente na meiose e diz-se que os alelos dos dois pares de genes heterozigotos demonstram **segregação (distribuição) independente**.

Este capítulo explica como podemos reconhecer a segregação independente e como tal princípio pode ser utilizado na construção de linhagens, tanto na agricultura quanto na pesquisa genética básica. (O Capítulo 4 abrange os princípios análogos aplicáveis aos pares de genes heterozigotos no *mesmo* par de cromossomos.)

Também devemos observar que a distribuição independente de uma variedade de genes também é útil para fornecer um mecanismo hereditário básico em relação aos fenótipos contínuos. Essas são propriedades tais como a estatura ou o peso, que não se encontram em categorias distintas, mas que, apesar disso, com frequência são influenciadas fortemente por genes múltiplos, coletivamente denominados "poligenes". Examinaremos o papel da distribuição independente na herança de fenótipos contínuos influenciados pelos referidos poligenes. Observaremos que a distribuição independente dos poligenes pode produzir uma distribuição fenotípica contínua entre a progênie.

Por último, apresentaremos um tipo diferente de herança independente, aquele dos genes nas organelas mitocôndrias e cloroplastos. Contrariamente aos cromossomos nucleares, esses genes são herdados citoplasmaticamente e resultam em padrões de herança diferentes daqueles observados em relação aos genes e cromossomos nucleares. Esse padrão é independente dos genes que demonstram

herança nuclear.

FIGURA 3.1 Genótipos superiores de cultivos tal como o arroz revolucionaram a agricultura. Esta fotografia demonstra alguns dos genótipos-chave utilizados em programas de cultivo de arroz. (Bloomberg/Getty Images.)

Primeiramente, examinamos os procedimentos analíticos que se referem à distribuição independente dos genes nucleares. Esses foram desenvolvidos pela primeira vez pelo pai da genética, Gregor Mendel. Assim, novamente voltamos para o seu trabalho como um exemplo prototípico.

3.1 Lei de Mendel de segregação independente

Em uma grande parte do seu trabalho original com ervilhas, Mendel analisou os descendentes de linhagens puras que diferiam em *duas* características. O simbolismo geral a seguir é utilizado para representar genótipos que incluem dois genes. Se dois genes estão em cromossomos diferentes, os pares de genes são separados por um ponto e vírgula — por exemplo, *A*/*a*; *B*/*b*. Se eles estão no mesmo cromossomo, os alelos em um homólogo são escritos de modo adjacente,

sem pontuação, e são separados daqueles no outro homólogo por uma barra — por exemplo, *AB*/*ab* ou *Ab*/*aB*. Não existe um simbolismo aceito para situações nas quais não se sabe se os genes estão no mesmo cromossomo ou em cromossomos diferentes. Para essa situação de posição desconhecida utilizaremos, neste livro, um ponto para separar os genes — por exemplo, *A*/*a* · *B*/*b*. Relembre, do Capítulo 2, que um heterozigoto para um *gene único* (tal como *A*/*a*) por vezes é denominado um mono-híbrido; um heterozigoto *duplo*, tal como *A*/*a* · *B*/*b* por vezes é denominado um **di-híbrido**. A partir do estudo de **cruzamentos di-híbridos** (*A*/*a* · *B*/*b* × *A*/*a* · *B*/*b*), Mendel inventou seu segundo princípio de hereditariedade importante, a **lei da segregação independente**, por vezes denominada **segunda lei de Mendel**.

O par de características com o qual ele começou a trabalhar foi o formato e a cor da semente. Nós já acompanhamos o cruzamento mono-híbrido em relação à cor da semente (*Y*/*y* × *Y*/*y*), que forneceu uma proporção da progênie de 3 amarelas:1 verde. Os fenótipos do formato da semente (Figura 3.2) eram lisa (determinado pelo alelo *R*) e rugosa (determinada pelo alelo *r*). O cruzamento mono-híbrido *R*/*r* × *R*/*r* forneceu uma proporção de progênie de 3 lisas:1 rugosa, conforme esperado (ver Tabela 2.1, no Capítulo 2). Para realizar um cruzamento di-híbrido, Mendel iniciou com duas linhagens parentais puras. Uma linhagem apresentava sementes rugosas e amarelas. Tendo em vista que Mendel não tinha o conceito da localização cromossômica dos genes, precisamos utilizar a representação com ponto para escrever o genótipo combinado inicialmente como *r*/*r* · *Y*/*Y*. A outra linhagem apresentava sementes lisas e verdes, com o genótipo *R*/*R* · *y*/*y*. Quando essas duas linhagens foram cruzadas, elas necessariamente produziram gametas que eram *r* · *Y* e *R* · *y*, respectivamente. Portanto, as sementes da F_1 tinham de ser di-híbridas, do genótipo *R*/*r* · *Y*/*y*. Mendel descobriu que as sementes da F_1 eram lisas e amarelas. Esse resultado demonstrou que a dominância de *R* sobre *r* e de *Y* sobre *y* não foi afetada pela condição do outro par de genes no di-híbrido *R*/*r* · *Y*/*y*. Em outras palavras, *R* permaneceu dominante sobre *r*, independentemente da cor da semente, e *Y* permaneceu dominante sobre *y*, independentemente do formato da semente.

FIGURA 3.2 Ervilhas lisas (*R*/*R* ou *R*/*r*) e rugosas (*r*/*r*) estão presentes em uma vagem de uma planta heterozigota (*R*/*r*) autocruzada. A proporção fenotípica nesta vagem é precisamente a proporção de 3:1 esperada em média na progênie deste autocruzamento. (Estudos moleculares demonstraram que o alelo rugoso utilizado por Mendel é produzido pela inserção de um segmento de DNA móvel dentro do gene; ver Capítulo 15.) (Madan K. Bhattacharyya.)

Em seguida, Mendel autocruzou a F_1 di-híbrida para obter a geração F_2. As sementes da F_2 eram de 4 diferentes tipos, nas proporções a seguir:

$\frac{9}{16}$ de sementes lisas e amarelas

$\frac{3}{16}$ de sementes lisas e verdes

$\frac{3}{16}$ de sementes rugosas e amarelas

$\frac{1}{16}$ de sementes rugosas e verdes

um resultado que está ilustrado na Figura 3.3 com os números reais obtidos por Mendel. Essa proporção, 9:3:3:1, inicialmente inesperada em relação a essas duas características, aparenta ser muito mais complexa do que as proporções de

3:1 simples dos cruzamentos mono-híbridos. Não obstante, a proporção de 9:3:3:1 comprovou ser um padrão de herança consistente em ervilhas. Como evidência, Mendel também realizou cruzamentos di-híbridos, que incluíram diversas outras combinações de características e observou que *todos* os indivíduos da F_1 di-híbridos produziram proporções de 9:3:3:1 na F_2. A proporção era outro padrão de herança que exigiu o desenvolvimento de uma nova ideia para a sua explicação.

Primeiramente, verificaremos os números reais obtidos por Mendel na Figura 3.3 para determinar se as proporções de 3:1 de mono-híbridos ainda podem ser observadas na F_2. Em relação ao formato da semente, existem 423 sementes lisas (315 + 108) e 133 sementes rugosas (101 + 32). Esse resultado está próximo de uma proporção de 3:1 (na realidade, 3,2:1). Em seguida, em relação à cor da semente, existem 416 sementes amarelas (315 + 101) e 140 verdes (108 + 32), também muito próximo de uma proporção de 3:1 (quase exatamente 3:1). A presença dessas duas proporções de 3:1 escondidas na proporção de 9:3:3:1 era indubitavelmente uma fonte da percepção que Mendel necessitava para explicar a proporção de 9:3:3:1, tendo em vista que ele percebeu que se tratava simplesmente de duas proporções de 3:1 diferentes, combinadas aleatoriamente. Um modo de visualizar a combinação aleatória dessas duas proporções é com um diagrama ramificado, como segue:

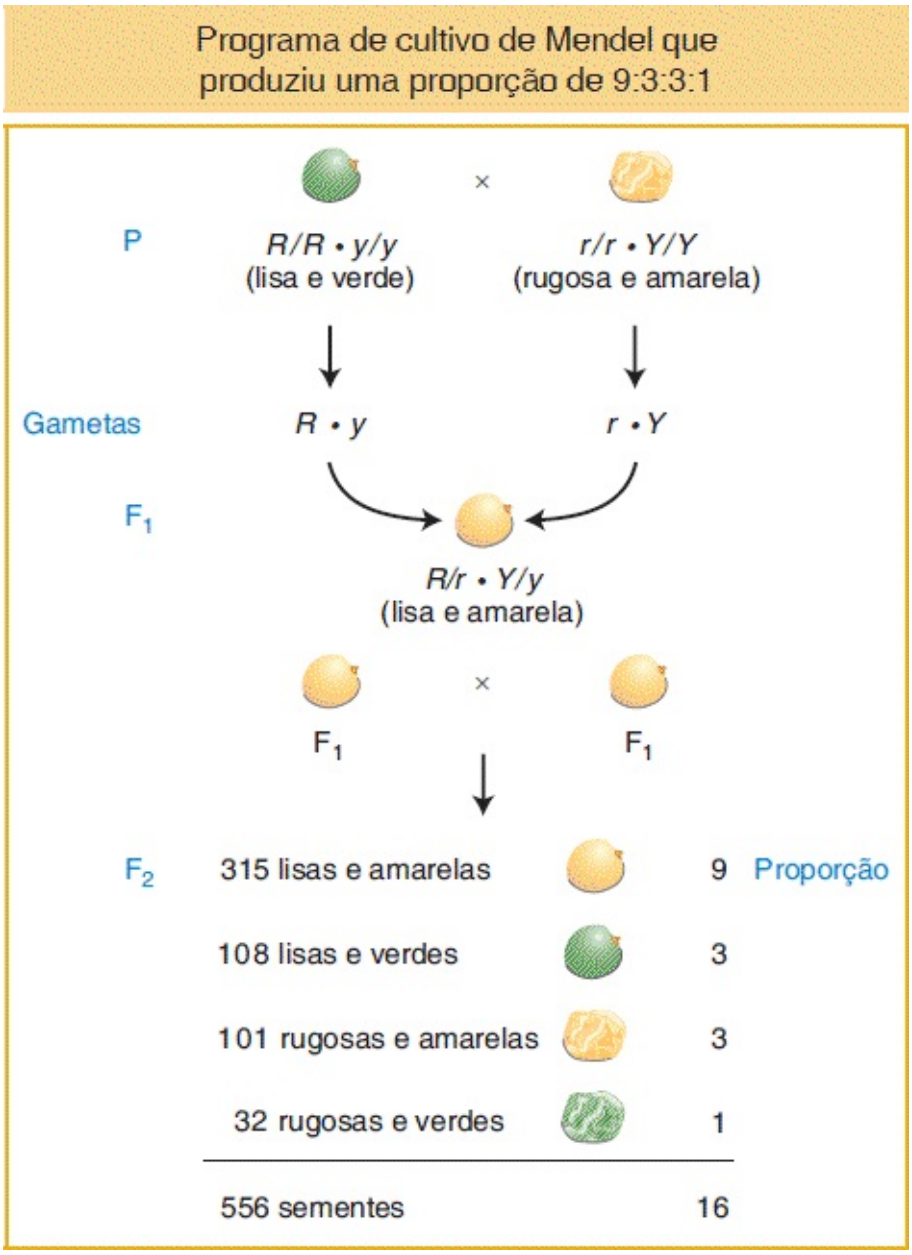
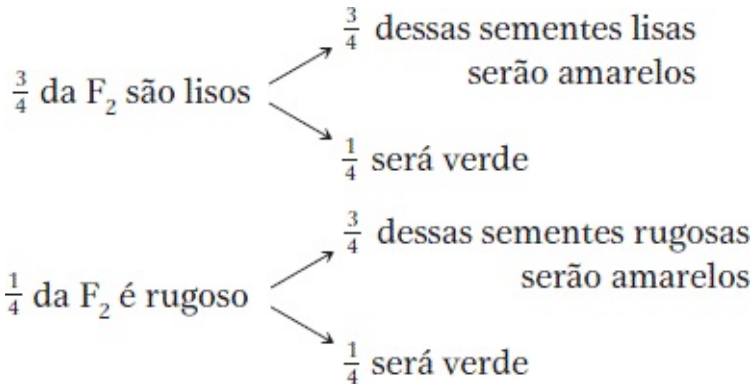


FIGURA 3.3 Mendel sintetizou um di-híbrido que, quando autocruzado, produziu uma progênie da F_2 na proporção de 9:3:3:1.



As probabilidades dos quatro resultados possíveis são calculadas por meio da utilização da regra do produto (a probabilidade de dois eventos independentes ocorrerem em conjunto é o produto de suas probabilidades individuais). Portanto, multiplicamos ao longo das ramificações no diagrama. Por exemplo, 3/4 de todas as sementes serão lisas e 3/4 das sementes lisas serão amarelas, de modo que a probabilidade de uma semente ser lisa e amarela é calculada como 3/4 × 3/4, que é igual a 9/16. Essas multiplicações fornecem as quatro proporções a seguir:

$$\frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \text{ lisas e amarelas}$$
$$\frac{3}{4} \times \frac{1}{4} = \frac{3}{16} \text{ lisas e verdes}$$
$$\frac{1}{4} \times \frac{3}{4} = \frac{3}{16} \text{ rugosas e amarelas}$$
$$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \text{ rugosa e verde}$$

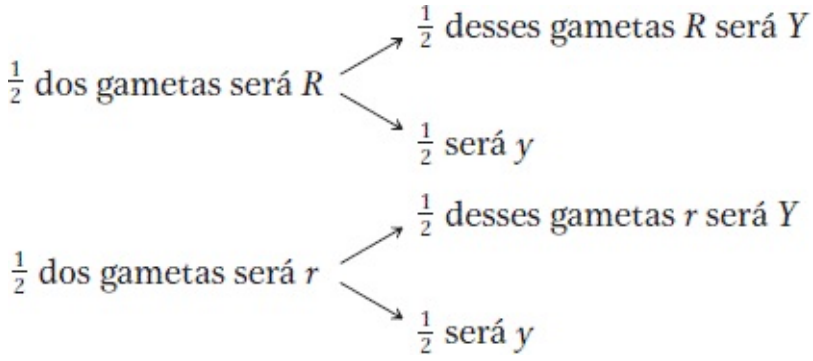
Essas proporções constituem a proporção de 9:3:3:1 que estamos tentando explicar. Entretanto, esse exercício não é meramente um malabarismo com números? O que a combinação das duas proporções de 3:1 significa biologicamente? O modo pelo qual Mendel escreveu sua explicação de fato equivale a um mecanismo biológico. No que atualmente é conhecido como a lei de segregação independente (segunda lei de Mendel), ele concluiu que *diferentes pares de genes se distribuem independentemente durante a formação dos gametas*. A consequência é que, em relação a dois pares de genes heterozigotos *A*/*a* e *B*/*b*, o alelo *b* apresenta justamente a mesma probabilidade de figurar em um gameta com um alelo *a* como com um alelo *A* e, da mesma maneira, em relação ao alelo *B*. Retrospectivamente, sabemos que, na maioria dos casos, essa “lei” se

aplica aos genes em cromossomos diferentes. Os genes no mesmo cromossomo em geral não se distribuem independentemente, tendo em vista que são mantidos em conjunto pelo próprio cromossomo.

> CONCEITO-CHAVE A segunda lei de Mendel (o princípio de segregação independente) declara que pares de genes em pares de cromossomos diferentes se distribuem independentemente na meiose.

A declaração original de Mendel dessa lei foi que genes diferentes se distribuem independentemente, tendo em vista que aparentemente ele não encontrou (ou ignorou) quaisquer exceções que pudessem ter levado ao conceito de ligação.

Explicamos a proporção fenotípica de 9:3:3:1 como duas proporções fenotípicas de 3:1 combinadas aleatoriamente. Mas também podemos chegar à proporção de 9:3:3:1 a partir de uma consideração da frequência dos gametas, os reais produtos meióticos? Consideremos os gametas produzidos pelo di-híbrido da F_1 R/r; Y/y (o ponto e vírgula demonstra que agora estamos abraçando a ideia de que os genes estão em cromossomos diferentes). Novamente, utilizaremos o diagrama ramificado para iniciarmos, tendo em vista que ele ilustra visualmente a independência. Combinando as leis de Mendel sobre a segregação igual e sobre a segregação independente, podemos prever que:



A multiplicação ao longo das ramificações de acordo com a regra do produto nos fornece as proporções de gametas:

$\frac{1}{4}$ *R*; *Y*

$\frac{1}{4}$ *R*; *y*

$\frac{1}{4}$ *r*; *Y*

$\frac{1}{4}$ *r*; *y*

Essas proporções são um resultado direto da aplicação das duas leis mendelianas: da segregação e da independência. Entretanto, ainda não chegamos à proporção de 9:3:3:1. A próxima etapa é reconhecer que, tendo em vista que os gametas masculinos e femininos obedecem às mesmas leis durante a formação, ambos os gametas masculinos e femininos demonstrarão as mesmas proporções há pouco fornecidas. Os quatro tipos gaméticos femininos serão fertilizados aleatoriamente pelos quatro tipos gaméticos masculinos para obter a F_2. O melhor modo gráfico de demonstrar os desfechos do cruzamento é por meio da utilização de uma grade de 4 × 4 denominada *quadrado de Punnett*, que está ilustrado na Figura 3.4. Já observamos que as grades são úteis em genética por permitirem uma representação visual dos dados. Sua utilidade está no fato de que as suas proporções podem ser desenhadas de acordo com as proporções genéticas ou as proporções em consideração. No quadrado de Punnett na Figura 3.4, por exemplo, foram desenhadas quatro linhas e quatro colunas para corresponder aos quatro genótipos dos gametas femininos e aos quatro dos gametas masculinos. Observamos que existem 16 quadros, que representam as diversas fusões gaméticas, e que cada quadro é 1/16 da área total da grade. De acordo com a regra do produto, cada 1/16 é um resultado da fertilização de um tipo de ovócito a uma frequência de 1/4 por um tipo de espermatozoide, também a uma frequência de 1/4, fornecendo a probabilidade daquela fusão como $(1/4)^2$. Conforme o quadrado de Punnett demonstra, a F_2 contém uma diversidade de genótipos, mas existem apenas quatro fenótipos e suas proporções estão na proporção de 9:3:3:1. Assim, observamos que quando calculamos as frequências da progênie diretamente por meio das frequências dos gametas, ainda chegamos à proporção de 9:3:3:1. Portanto, a lei de Mendel explica não apenas os fenótipos da F_2, mas também os genótipos dos gametas e da progênie que são a base da proporção

fenotípica da F_2.

Mendel prosseguiu para testar o seu princípio de segregação independente de diversos modos. O modo mais direto enfocou na proporção gamética de 1:1:1:1 supostamente produzida pelo di-híbrido da F_1 *R*/*r*; *Y*/*y*, tendo em vista que essa proporção surgiu diretamente de seu princípio da distribuição independente e foi a base biológica da proporção de 9:3:3:1 na F_2, conforme demonstrado pelo quadrado de Punnett. Para verificar a proporção gamética de 1:1:1:1, Mendel utilizou um cruzamento-teste. Ele realizou o cruzamento-teste do di-híbrido da F_1 com um testador de genótipo *r*/*r*; *y*/*y*, que produz apenas gametas com alelos recessivos (genótipo *r*; *y*). Ele ponderou que, se de fato havia uma proporção de 1:1:1:1 de gametas *R*; *Y*, *R*; *y*, *r*; *Y* e *r*; *y*, as proporções da progênie desse cruzamento deveriam corresponder diretamente às proporções gaméticas produzidas pelo di-híbrido. Em outras palavras:

Quadrado de Punnett ilustrando os genótipos subjacentes a uma proporção de 9:3:3:1

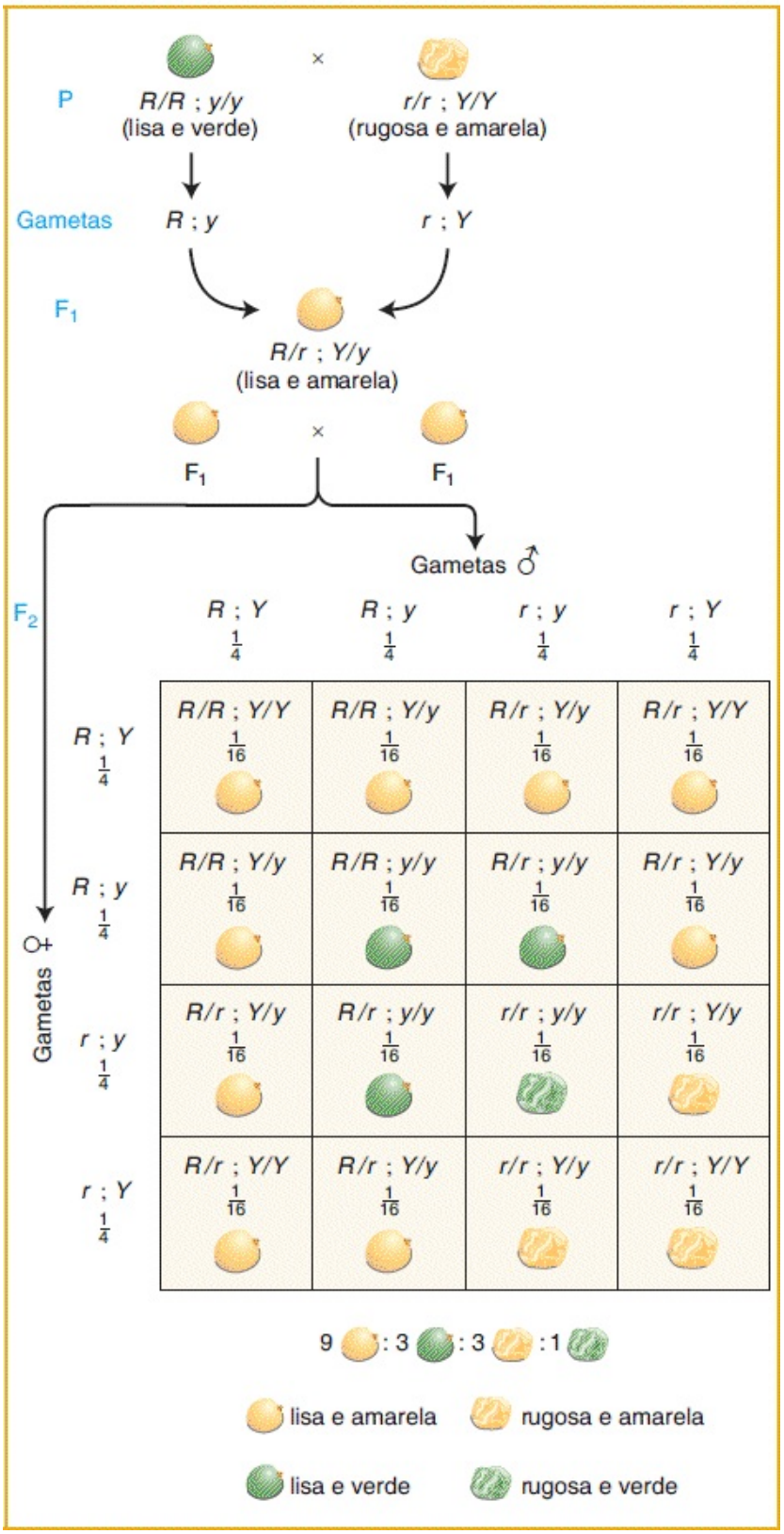


FIGURA 3.4 Podemos utilizar o quadrado de Punnett para prever o resultado de um cruzamento di-híbrido. Este quadrado demonstra a constituição genotípica e fenotípica prevista na geração F_2 a partir de um cruzamento di-híbrido.

$\frac{1}{4}$ *R*/*r*; *Y*/*y* → lisas e amarelas

$\frac{1}{4}$ *R*/*r*; *y*/*y* → lisas e verdes

$\frac{1}{4}$ *r*/*r*; *Y*/*y* → rugosas e amarelas

$\frac{1}{4}$ *r*/*r*; *y*/*y* → rugosas e verdes

Essas proporções foram o resultado que ele obteve, perfeitamente consistente com as suas expectativas. Ele obteve resultados semelhantes em relação a todos os outros cruzamentos di-híbridos que realizou e tais testes e outros tipos de testes demonstraram, todos, que de fato ele havia planejado um modelo robusto para explicar os padrões de herança observados em seus diversos cruzamentos de ervilhas.

No início do século 20, ambas as leis de Mendel foram testadas em um amplo espectro de organismos eucarióticos. Os resultados desses testes demonstraram que os princípios mendelianos em geral eram aplicáveis. As proporções mendelianas (tais como 3:1, 1:1, 9:3:3:1 e 1:1:1:1) foram extensivamente relatadas, sugerindo que a segregação igual e a distribuição independente são processos hereditários fundamentais observados em toda a natureza. As leis de Mendel não são meramente leis a respeito das ervilhas. São leis a respeito da genética dos organismos eucarióticos em geral.

Como um exemplo da aplicabilidade universal do princípio da segregação independente, podemos examinar a sua ação em haploides. Se o princípio da segregação igual é válido de modo geral, então devemos ser capazes de observar a sua ação em haploides, tendo em vista que os haploides são submetidos à meiose. De fato, a distribuição independente pode ser observada em um cruzamento do tipo *A*; *B* × *a*; *b*. A fusão das células parentais resulta em um meiócito diploide temporário que é um di-híbrido *A*/*a*; *B*/*b* e os produtos da meiose aleatoriamente amostrados (esporos sexuados, tais como ascósporos em

fungos) serão:

$$\frac{1}{4}A; B$$

$$\frac{1}{4}A; b$$

$$\frac{1}{4}a; B$$

$$\frac{1}{4}a; b$$

Portanto, observamos a mesma proporção do cruzamento-teste di-híbrido em um organismo diploide; novamente, a proporção é uma combinação aleatória das duas proporções de 1:1 mono-híbridas, em virtude da distribuição independente.

> CONCEITO-CHAVE As proporções de 1:1:1:1 e 9:3:3:1 são diagnósticas da distribuição independente em um e dois meiócitos di-híbridos, respectivamente.

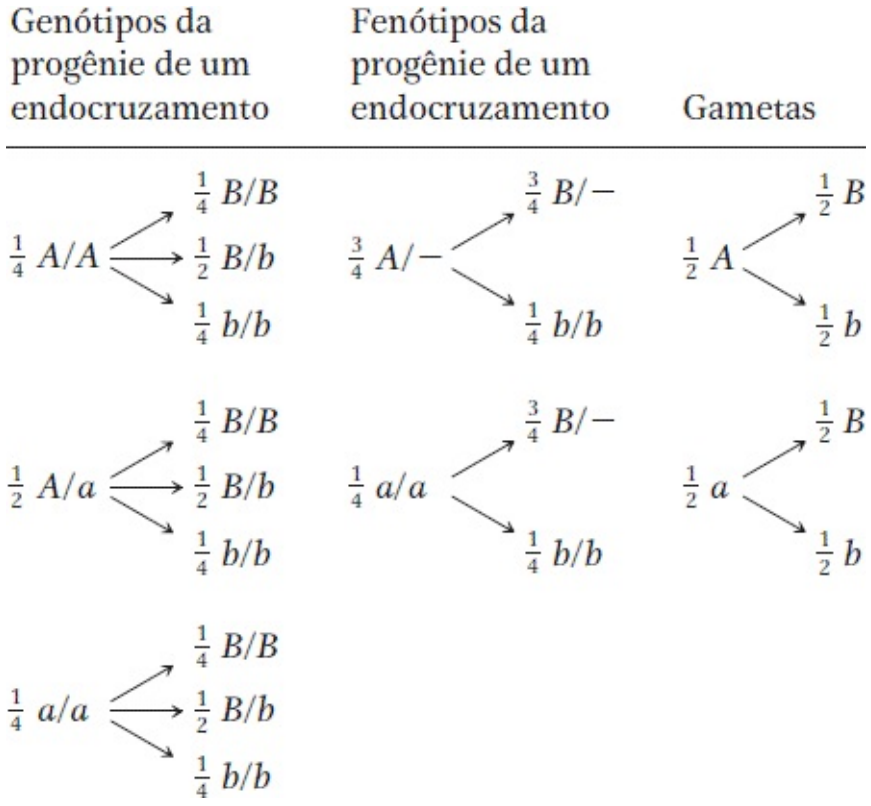
3.2 A segregação independente

Nesta seção, examinaremos diversos procedimentos analíticos que são parte da pesquisa genética diária e todos têm por base o conceito da distribuição independente. Todos esses procedimentos são utilizados para analisar as proporções fenotípicas.

Previsão das proporções da progênie

A genética pode atuar em qualquer uma de duas direções: (1) prevendo os genótipos desconhecidos dos progenitores por meio da utilização das proporções fenotípicas da progênie ou (2) prevendo as proporções fenotípicas da progênie a partir dos progenitores de genótipo conhecido. A última é uma parte importante da genética pertinente à previsão dos tipos de progênie que surgirão a partir de um cruzamento e ao cálculo de suas frequências esperadas — em outras palavras, as suas probabilidades. Isso é útil não apenas na pesquisa em organismos-modelo,

mas também para prever os desfechos de cruzamentos na genética humana; por exemplo, no aconselhamento genético, pessoas apreciam estimativas de risco específicas. Já examinamos dois métodos para a previsão: os quadrados de Punnett e os diagramas ramificados. Os quadrados de Punnett podem ser utilizados para demonstrar padrões hereditários com base em um par de genes, dois pares de genes ou mais. As referidas grades são bons dispositivos gráficos para a representação da progênie, mas desenhá-los demanda tempo. Até mesmo o quadrado de Punnett de 16 compartimentos que utilizamos para analisar um cruzamento di-híbrido demora um longo tempo para ser escrito, mas em relação a um cruzamento tri-híbrido, existem 2^3, ou 8 tipos de gametas diferentes, e o quadrado de Punnett apresenta 64 compartimentos. O diagrama ramificado (demonstrado a seguir) é de mais fácil criação e é adaptável às proporções fenotípicas, genotípicas ou gaméticas, conforme ilustrado em relação ao di-híbrido *A*/*a*; *B*/*b*.



Observe, entretanto, que a “árvore” de ramificações em relação aos genótipos é um tanto complicada até mesmo nesse caso simples, que utiliza dois pares de

genes, tendo em vista que existem $3^2 = 9$ genótipos. Em relação a três pares de genes, existem 3^3, ou 27 possíveis genótipos. Para simplificar esse problema, podemos utilizar uma abordagem estatística, que constitui um terceiro método para o cálculo das probabilidades (frequências esperadas) de fenótipos ou genótipos específicos resultantes de um cruzamento. As duas regras estatísticas necessárias são a **regra do produto** (introduzida no Capítulo 2) e a **regra da soma**, que agora consideramos em conjunto.

> CONCEITO-CHAVE A regra do produto estabelece que a probabilidade de eventos independentes ocorrerem em conjunto é o produto de suas probabilidades individuais.

Os desfechos possíveis ao jogar dois dados seguem a regra do produto, tendo em vista que o resultado em um dado é independente do outro. Como um exemplo, calculemos a probabilidade, *p*, de obtermos um par de 4. A probabilidade de um 4 em um dado é de 1/6, tendo em vista que o dado apresenta seis lados e apenas um lado contém o número 4. A probabilidade é escrita como segue:

$$p\ (\text{um } 4) = \frac{1}{6}$$

Portanto, com a utilização da regra do produto, a probabilidade de um 4 aparecer em ambos os dados é de 1/6 × 1/6 = 1/36, que é escrita:

$$p\ (\text{dois } 4) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Agora, em relação à regra da soma:

> CONCEITO-CHAVE A regra da soma estabelece que a probabilidade de *um* ou de *outro* de dois eventos mutuamente exclusivos ocorrer é a soma das suas probabilidades individuais.

(Observe que, na regra do produto, o enfoque está nos desfechos A *e* B. Na

regra da soma, o enfoque está no desfecho A' ou A''.)

Dados também podem ser utilizados para ilustrar a regra da soma. Já calculamos que a probabilidade de dois 4 é de 1/36; claramente, com a utilização do mesmo tipo de cálculo, a probabilidade de dois 5 será a mesma ou 1/36. Agora, podemos calcular a probabilidade de dois 4 *ou* de dois 5. Tendo em vista que esses desfechos são mutuamente exclusivos, a regra da soma pode ser utilizada para nos informar que a resposta é 1/36 + 1/36, que é 1/18. A probabilidade pode ser escrita como segue:

$$p \text{ (dois 4 ou dois 5)} = \frac{1}{36} + \frac{1}{36} = \frac{1}{18}$$

Qual proporção da progênie será de um genótipo específico? Agora podemos nos voltar para um exemplo genético. Presuma que temos duas plantas de genótipos:

A/*a*; *b*/*b*; *C*/*c*; *D*/*d*; *E*/*e*

e

A/*a*; *B*/*b*; *C*/*c*; *d*/*d*; *E*/*e*

A partir de um cruzamento entre essas plantas, desejamos recuperar uma planta da progênie de genótipo *a*/*a*; *b*/*b*; *c*/*c*; *d*/*d*; *e*/*e* (talvez com a finalidade de atuar como a linhagem testadora em um cruzamento-teste). Qual proporção da progênie devemos esperar que seja daquele genótipo? Se presumirmos que todos os pares de genes se distribuem independentemente, então podemos realizar esse cálculo facilmente por meio da utilização da regra do produto. Os cinco pares de genes diferentes são considerados individualmente, como se fossem cinco cruzamentos separados e em seguida as probabilidades individuais de obtenção de cada genótipo são multiplicadas em conjunto para chegar à resposta:

De *A*/*a* × *A*/*a*, um quarto da progênie será *a*/*a*.

De *b*/*b* × *B*/*b*, metade da progênie será *b*/*b*.

De *C*/*c* × *C*/*c*, um quarto da progênie será *c*/*c*.

De *D*/*d* × *d*/*d*, metade da progênie será *d*/*d*.

De *E*/*e* × *E*/*e*, um quarto da progênie será *e*/*e*.

Portanto, a probabilidade geral (ou a frequência esperada) de obtenção de uma progênie de genótipo *a*/*a*; *b*/*b*; *c*/*c*; *d*/*d*; *e*/*e* será 1/4 × 1/2 × 1/4 × 1/2 × 1/4 = 1/256. Esse cálculo da probabilidade pode ser estendido para prever as frequências fenotípicas ou as frequências gaméticas. De fato, existem muitas outras utilizações para esse método na análise genética e encontraremos algumas nos capítulos posteriores.

Quantos descendentes precisamos cultivar? Para adotar o exemplo antecedente em uma etapa adiante, suponha que precisemos estimar quantas plantas da progênie precisam ser cultivadas para haver uma chance razoável de obtenção do genótipo desejado *a*/*a*; *b*/*b*; *c*/*c*; *d*/*d*; *e*/*e*. Primeiramente, calculamos a proporção da progênie que se espera ser daquele genótipo. Conforme demonstrado há pouco, aprendemos que precisamos examinar no mínimo 256 progênies para haver uma chance média de obtenção de uma planta individual do genótipo desejado.

A probabilidade de obtenção de um "sucesso" (uma planta totalmente recessiva) entre 256 precisa ser considerada mais cuidadosamente. Essa é a probabilidade *média* de sucesso. Infelizmente, se isolarmos e testarmos 256 progênies, muito provavelmente não obteremos sucesso em absoluto, simplesmente em virtude de azar. De um ponto de vista prático, uma questão mais significativa a ser indagada seria: "Qual tamanho de amostra necessitamos para ter uma confiança de 95% de que obteremos no mínimo um sucesso?" (Nota: esse valor de confiança de 95% é o padrão em ciência.) O modo mais simples de realizar esse cálculo é abordá-lo ao considerar a probabilidade de completo insucesso — ou seja, a probabilidade de não obtermos nenhum indivíduo do genótipo desejado. Em nosso exemplo, para cada indivíduo isolado, a probabilidade de que ele *não* seja do tipo desejado é 1 — (1/256) = 255/256. Estendendo essa ideia para uma amostra de tamanho *n*, observamos que a probabilidade de nenhum sucesso em uma amostra de *n* é $(255/256)^n$. (Essa probabilidade é uma aplicação simples da regra do produto: 255/256 multiplicada por si própria *n* vezes.) Portanto, a probabilidade de obtenção de *no*

mínimo um sucesso é a probabilidade de todos os desfechos possíveis (essa probabilidade é 1) menos a probabilidade de insucesso total ou $(255/256)^n$. Portanto, a probabilidade de no mínimo um sucesso é $1 — (255/256)^n$. Para atender o nível de confiança de 95%, precisamos tornar essa expressão igual a 0,95 (o equivalente a 95%).

Portanto:

$$1 — (255/256)^n = 0{,}95$$

A solução dessa equação em relação ao *n* nos fornece um valor de 765, o número necessário de progênie para virtualmente garantir o sucesso. Observe quão diferente esse número é da expectativa ingênua de sucesso na progênie de 256. Esse tipo de cálculo é útil em muitas aplicações em genética e em outras situações nas quais é necessário um desfecho de sucesso de muitos estudos.

Quantos genótipos distintos um cruzamento produzirá? As regras da probabilidade podem ser facilmente utilizadas para prever a quantidade de genótipos ou fenótipos na progênie de linhagens parentais complexas. (Os referidos cálculos são utilizados de modo rotineiro em pesquisas, na análise de progênie e na criação de linhagens.) Por exemplo, em um endocruzamento do "tetra-híbrido" *A*/*a*; *B*/*b*; *C*/*c*; *D*/*d*, haverá três genótipos para cada par de genes; por exemplo, em relação ao primeiro par de genes, os três genótipos serão *A*/*a*, *A*/*A* e *a*/*a*. Tendo em vista que existem quatro pares de genes no total, haverá $3^4 = 81$ genótipos diferentes. Em um cruzamento-teste de um referido tetra-híbrido, haverá dois genótipos para cada par de genes (p. ex., *A*/*a* e *a*/*a*) e um total de $2^4 = 16$ genótipos na progênie. Tendo em vista que estamos presumindo que todos os genes se encontram em cromossomos diferentes, todos esses genótipos do cruzamento-teste ocorrerão a uma frequência igual de 1/16.

Utilização do teste do qui-quadrado em proporções mono-híbridas e di-híbridas

Na genética em geral, um pesquisador com frequência é confrontado com

resultados que estão próximos de uma proporção esperada, mas que não são idênticos a ela. As referidas proporções podem ser de mono-híbridos, di-híbridos, ou genótipos mais complexos e com independência ou não. Mas quão próximo de um resultado esperado é suficiente? É necessário um teste estatístico para verificar os referidos números em face das expectativas e o **teste do qui-quadrado**, ou teste do χ^2, preenche esse papel.

Em quais situações experimentais o teste de χ^2 em geral é aplicável? A situação geral é uma situação na qual os resultados observados são comparados àqueles previstos por uma hipótese. Em um exemplo genético simples, suponha que você cultivou uma planta em relação à qual você formula a hipótese de que, com base em uma análise precedente, seja uma heterozigota, *A*/*a*. Para testar essa hipótese, você cruza essa heterozigota com um testador de genótipo *a*/*a* e conta os números de fenótipos com genótipos *A*/— e *a*/*a* na progênie. Em seguida, você deve avaliar se os números que você obtém constituem a proporção esperada de 1:1. Se houver uma correspondência próxima, então a hipótese é considerada consistente com o resultado, enquanto se houver uma correspondência inadequada, a hipótese é rejeitada. Como parte desse processo, deve ser feita uma avaliação se os números observados estão *suficientemente* próximos daqueles esperados. Correspondências muito próximas e incompatibilidades óbvias em geral não apresentam problemas, mas, inevitavelmente, existem áreas sombrias nas quais a correspondência não é óbvia.

O teste do χ^2 é simplesmente um modo de quantificar os diversos desvios aleatórios esperados se uma hipótese for verdadeira. Adote a hipótese simples precedente que prevê uma proporção de 1:1, por exemplo. Ainda que a hipótese seja verdadeira, apenas raramente podemos esperar uma proporção exata de 1:1. Podemos modelar essa ideia com um barril cheio de bolas de gude vermelhas e brancas em quantidades iguais. Se removermos amostras de 100 bolas de gude às cegas, com base no acaso esperaríamos que as amostras demonstrassem pequenos desvios, tais como 52 vermelhas:48 brancas, de modo consideravelmente comum e que demonstrassem desvios maiores, tais como 60 vermelhas:40 brancas menos comumente. Até mesmo 100 bolas de gude vermelhas é um desfecho possível, a uma probabilidade muito baixa de $(1/2)^{100}$. Entretanto, se *qualquer* resultado for

possível em algum nível de probabilidade, até mesmo se a hipótese for verdadeira, como podemos chegar a rejeitar uma hipótese? Uma convenção científica geral é que uma hipótese será rejeitada como falsa se houver uma probabilidade inferior a 5% de observação de um desvio das expectativas no mínimo tão grande quanto aquele realmente observado. A hipótese pode ainda ser verdadeira, mas precisamos tomar uma decisão em algum ponto e 5% é a linha de decisão convencional. A implicação é que, embora os resultados assim tão distantes das expectativas sejam esperados em 5% das ocasiões, até mesmo quando a hipótese é verdadeira, rejeitaremos erroneamente a hipótese em apenas 5% dos casos, e estamos desejando aceitar essa chance de erro. (Esses 5% são o contrário do nível de confiança de 95% utilizado anteriormente.)

Vejamos alguns dados reais. Testaremos a nossa hipótese anterior de que uma planta seja uma heterozigota. Consideraremos *A* para pétalas vermelhas e consideraremos *a* para brancas. Os cientistas testam uma hipótese ao realizar previsões com base na hipótese. Na presente situação, uma possibilidade é prever os resultados de um cruzamento-teste. Presuma que nós realizamos o cruzamento-teste do heterozigoto presumido. Com base na hipótese, a lei de Mendel da segregação igual prevê que deveremos ter 50% *A*/*a* e 50% *a*/*a*. Presuma que, na realidade, obtemos uma progênie de 120 e observamos que 55 são vermelhas e 65 são brancas. Esses números diferem das expectativas precisas, que seriam 60 vermelhas e 60 brancas. O resultado aparenta estar um pouco longe da proporção esperada, o que gera incerteza; assim, precisamos utilizar o teste do χ^2. Calculamos o χ^2 por meio da utilização da fórmula a seguir:

$$\chi^2 = \Sigma\ (O - E)^2/E \text{ para todas as classes}$$

na qual *E* é o número esperado em uma classe, *O* é o número observado em uma classe, e Σ significa “soma de”. O valor resultante, χ^2, fornecerá um valor numérico que estima o grau de concordância entre os resultados esperados (hipotetizados) e observados (reais), com o número crescendo ainda mais na medida em que a concordância aumenta.

O cálculo é mais simplesmente realizado por meio da utilização de uma tabela:

Classe	*O*	*E*	$(O - E)^2$	$(O - E)^2/E$
Vermelha	55	60	25	25/60 = 0,42
Branca	65	60	25	25/60 = 0,42
				Total = χ^2 = 0,84

Agora precisamos procurar esse valor de χ^2 na Tabela 3.1, que nos fornecerá o valor da probabilidade que desejamos. As linhas na Tabela 3.1 listam diferentes valores de *graus de liberdade* (*gl*). O número de graus de liberdade é o número de variáveis independentes nos dados. No presente contexto, o número de variáveis independentes é simplesmente o número de classes fenotípicas menos 1. Nesse caso, gl = 2 — 1 = 1. Assim, observamos apenas a linha de 1 gl. Observamos que nosso valor de χ^2 de 0,84 encontra-se em algum ponto entre as colunas marcadas 0,5 e 0,1 — em outras palavras, entre 50% e 10%. Esse valor de probabilidade é muito superior ao valor limite de 5% e, assim, aceitamos os resultados observados como sendo compatíveis com a hipótese.

Seguem-se algumas observações importantes sobre a aplicação desse teste:

1. O que realmente significa o valor da probabilidade? É a probabilidade de observação de um desvio dos resultados esperados *no mínimo tão grande* (não *exatamente* esse desvio) com base no acaso se a hipótese estiver correta.
2. O fato de que os nossos resultados “passaram” no teste do qui-quadrado em virtude de $p > 0,05$ não significa que a hipótese seja verdadeira. Isso significa meramente que os resultados são compatíveis com aquela hipótese.

Entretanto, se houvéssemos obtido um valor de $p < 0,05$, teríamos sido forçados a rejeitar a hipótese. A ciência se refere totalmente a hipóteses que podem ser falsas, não trata da "verdade".

3. Devemos ser cuidadosos ao formular a hipótese, tendo em vista que presunções tácitas com frequência encontram-se encerradas dentro dela. A presente hipótese é um caso assim; se fôssemos declará-la cuidadosamente, deveríamos dizer que "o indivíduo em teste é um heterozigoto *A*/*a*, esses alelos demonstram segregação igual na meiose e as progênies *A*/*a* e *a*/*a* são de viabilidade igual". Investigaremos os efeitos do alelo sobre a viabilidade no Capítulo 6, mas, por enquanto, devemos tê-los em mente como uma possível complicação, tendo em vista que as diferenças na sobrevivência afetariam os tamanhos das diversas classes. O problema é que, se rejeitarmos uma hipótese que apresenta componentes escondidos, não sabemos quais dos componentes estamos rejeitando. Por exemplo, no presente caso, se fôssemos forçados a rejeitar a hipótese como resultado do teste do χ^2, não saberíamos se estamos rejeitando a segregação igual, ou a viabilidade igual, ou ambas.
4. O desfecho do teste do χ^2 depende fortemente dos tamanhos das amostras (números nas classes). Portanto, o teste deve utilizar *números reais*, não proporções ou porcentagens. Além disso, quanto maiores as amostras, mas poderoso é o teste.

Quaisquer proporções mendelianas familiares consideradas neste capítulo ou no Capítulo 2 podem ser testadas por meio da utilização do teste do χ^2 — por exemplo, 3:1 (1 gl), 1:2:1 (2 gl), 9:3:3:1 (3 gl) e 1:1:1:1 (3 gl). Retornaremos a mais aplicações do teste do χ^2 no Capítulo 4.

Tabela 3.1 Valores críticos da distribuição do χ^2.

					P		
gl	**0,995**	**0,975**	**0,9**	**0,5**	**0,1**	**0,05**	**0,025**

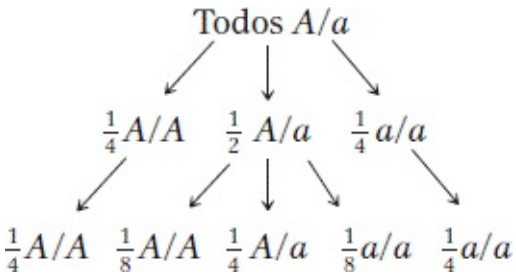
1	0,000	0,000	0,016	0,455	2,706	3,841	5,024
2	0,010	0,051	0,211	1,386	4,605	5,991	7,378
3	0,072	0,216	0,584	2,366	6,251	7,815	9,348
4	0,207	0,484	1,064	3,357	7,779	9,488	11,143
5	0,412	0,831	1,610	4,351	9,236	11,070	12,832
6	0,676	1,237	2,204	5,348	10,645	12,592	14,449
7	0,989	1,690	2,833	6,346	12,017	14,067	16,013
8	1,344	2,180	3,490	7,344	13,362	15,507	17,535
9	1,735	2,700	4,168	8,343	14,684	16,919	19,023
10	2,156	3,247	4,865	9,342	15,987	18,307	20,483
11	2,603	3,816	5,578	10,341	17,275	19,675	21,920
12	3,074	4,404	6,304	11,340	18,549	21,026	23,337
13	3,565	5,009	7,042	12,340	19,812	22,362	24,736

14	4,075	5,629	7,790	13,339	21,064	23,685	26,119
15	4,601	6,262	8,547	14,339	22,307	24,996	27,488

Síntese de linhagens puras

As linhagens puras estão entre as ferramentas essenciais da genética. Por um lado, apenas essas linhagens totalmente homozigotas expressarão alelos recessivos, mas a principal necessidade de linhagens puras está na manutenção de estoques para pesquisas. Os membros de uma linhagem pura podem ser entrecruzados ao longo do tempo e, assim, atuar como uma fonte constante do genótipo para utilização em experimentos. Portanto, em relação à maior parte dos organismos-modelo, existem centros de estoque internacionais que são repositórios das linhagens puras para utilização em pesquisas. Centros de estoque semelhantes fornecem linhagens de plantas e animais para utilização na agricultura.

As linhagens puras de plantas ou animais são produzidas por meio de gerações repetidas de autocruzamento. (Em animais, isso é feito por meio do cruzamento de animais de genótipo idêntico.) O autocruzamento de uma planta mono-híbrida demonstra o princípio em ação. Suponha que iniciemos com uma população de indivíduos que são todos *A*/*a* e que deixemos que ocorra autocruzamento. Podemos aplicar a primeira lei de Mendel para prever que, na próxima geração, haverá *A*/*A*, *A*/*a* e *a*/*a*. Observe que a *heterozigosidade* (a proporção de heterozigotos) foi dividida pela metade, de 1 para ½. Se repetirmos esse processo por outra geração, todos os descendentes dos homozigotos serão homozigotos, mas, novamente, os heterozigotos dividirão pela metade a sua proporção, para um quarto. O processo está demonstrado na representação a seguir:

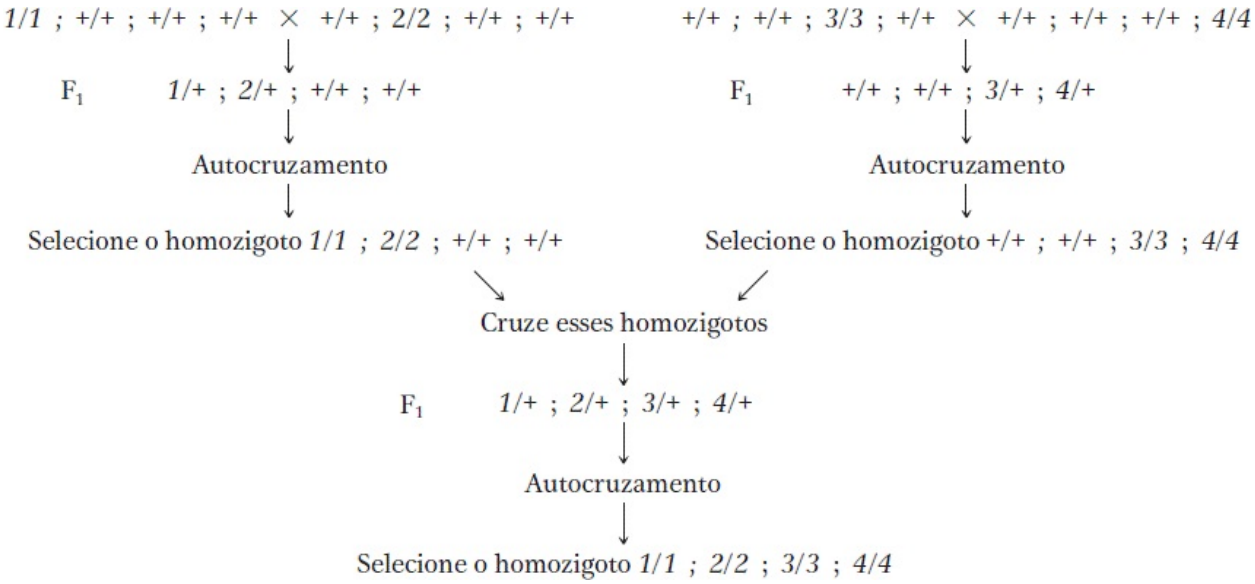


Após, digamos, 8 gerações de autocruzamento, a proporção de heterozigotos é reduzida para $(1/2)^8$, que é 1/256, ou aproximadamente 0,4%. Vejamos esse processo de um modo discretamente diferente: presumiremos que iniciamos um referido programa com um genótipo que é heterozigoto em 256 pares de genes. Se também presumirmos a distribuição independente, então, após o autocruzamento por oito gerações, terminaríamos com uma variedade de genótipos, cada um apresentando em média apenas um gene heterozigoto (ou seja, 1/256). Em outras palavras, estamos bem no caminho de criar um número de linhagens puras.

Apliquemos esse princípio para a seleção de linhagens agrícolas, o tópico com o qual iniciamos o capítulo. Podemos utilizar como nosso exemplo a seleção do trigo Marquis por Charles Saunders na parte inicial do século 20. O objetivo de Saunders era desenvolver uma linhagem de trigo produtiva, que apresentaria uma estação de crescimento mais curta e que, portanto, abriria grandes áreas de terreno em países do norte, tais como Canadá e Rússia, para o cultivo do trigo, outro dos alimentos básicos mundiais. Ele cruzou uma linhagem que apresentava excelente qualidade de grãos, denominada Red Fife, com uma linhagem denominada Hard Red Calcutta, a qual, embora de inadequada produção e qualidade, maturava 20 dias antes do que a Red Fife. A F_1 produzida pelo cruzamento presumivelmente era heterozigota para genes múltiplos que controlam as qualidades do trigo. A partir dessa F_1, Saunders realizou autocruzamentos e seleções que finalmente levaram a uma linhagem pura, que apresentava a combinação das propriedades favoráveis necessárias — grãos de boa qualidade e maturação precoce. Essa linhagem foi denominada Marquis. Ela foi rapidamente adotada em muitas partes do mundo.

Uma abordagem semelhante pode ser aplicada às linhagens de arroz com as quais iniciamos o capítulo. Todas as mutações monogênicas são cruzadas em

pares e, em seguida, as plantas das suas F_1 são auto- ou intercruzadas com outras plantas da F_1. Como uma demonstração, consideremos apenas quatro mutações, *1* a *4*. Um programa de cruzamento pode ser como segue, no qual os alelos mutantes e seus correspondentes do tipo selvagem sempre estão listados na mesma ordem (relembre que o sinal de + designa o tipo selvagem):



Esse tipo de cruzamento foi aplicado a muitas outras espécies de cultivo. As linhagens puras coloridas e diversas de tomates utilizadas no comércio estão demonstradas na Figura 3.5.

Observe que, em geral, quando um heterozigoto múltiplo é autocruzado, é produzida uma diversidade de diferentes homozigotos. Por exemplo, de *A*/*a*; *B*/*b*; *C*/*c*, existem dois homozigotos para cada par de genes (ou seja, em relação ao primeiro gene, os homozigotos são *A*/*A* e *a*/*a*) e, assim, existem $2^3 = 8$ homozigotos diferentes possíveis: *A*/*A*; *b*/*b*; *C*/*c* e *a*/*a*; *B*/*B*; *c*/*c*, e assim por diante. Cada homozigoto distinto pode ser o início de uma nova linhagem pura.

FIGURA 3.5 O cruzamento de tomates resultou em uma ampla diversidade de linhagens de diferentes genótipos e fenótipos. (©Mascarucci/Corbis.)

> CONCEITO-CHAVE O autocruzamento repetido leva a um aumento na proporção de homozigotos, um processo que pode ser utilizado para criar linhagens puras para pesquisas ou outras aplicações.

Vigor híbrido

Temos considerado a síntese das linhagens puras superiores para pesquisas e para a agricultura. As linhagens puras são convenientes, no sentido em que a propagação do genótipo de um ano para o outro é razoavelmente fácil. Entretanto, uma grande proporção de sementes comerciais que fazendeiros (e jardineiros) utilizam é denominada *semente híbrida*. Curiosamente, em muitos casos em que duas linhagens de plantas (e animais) discrepantes são unidas em um híbrido da F_1 (presumidamente heterozigoto), o híbrido demonstra tamanho e vigor

superiores aos das duas linhagens contribuintes (Figura 3.6). Essa superioridade geral dos heterozigotos múltiplos é denominada **vigor híbrido**. Os motivos moleculares para o vigor híbrido são, em sua maior parte, desconhecidos e ainda debatidos acaloradamente, mas o fenômeno é inegável e realizou grandes contribuições para a agricultura. Um aspecto negativo da utilização de híbridos é que, a cada estação, as duas linhagens parentais devem ser cultivadas em separado e em seguida devem ser entrecruzadas para a produção das sementes híbridas para a venda. Esse processo é muito mais inconveniente do que a manutenção de linhagens puras, que requer apenas o autocruzamento das plantas; consequentemente, a semente híbrida é mais cara do que a semente de linhagens puras.

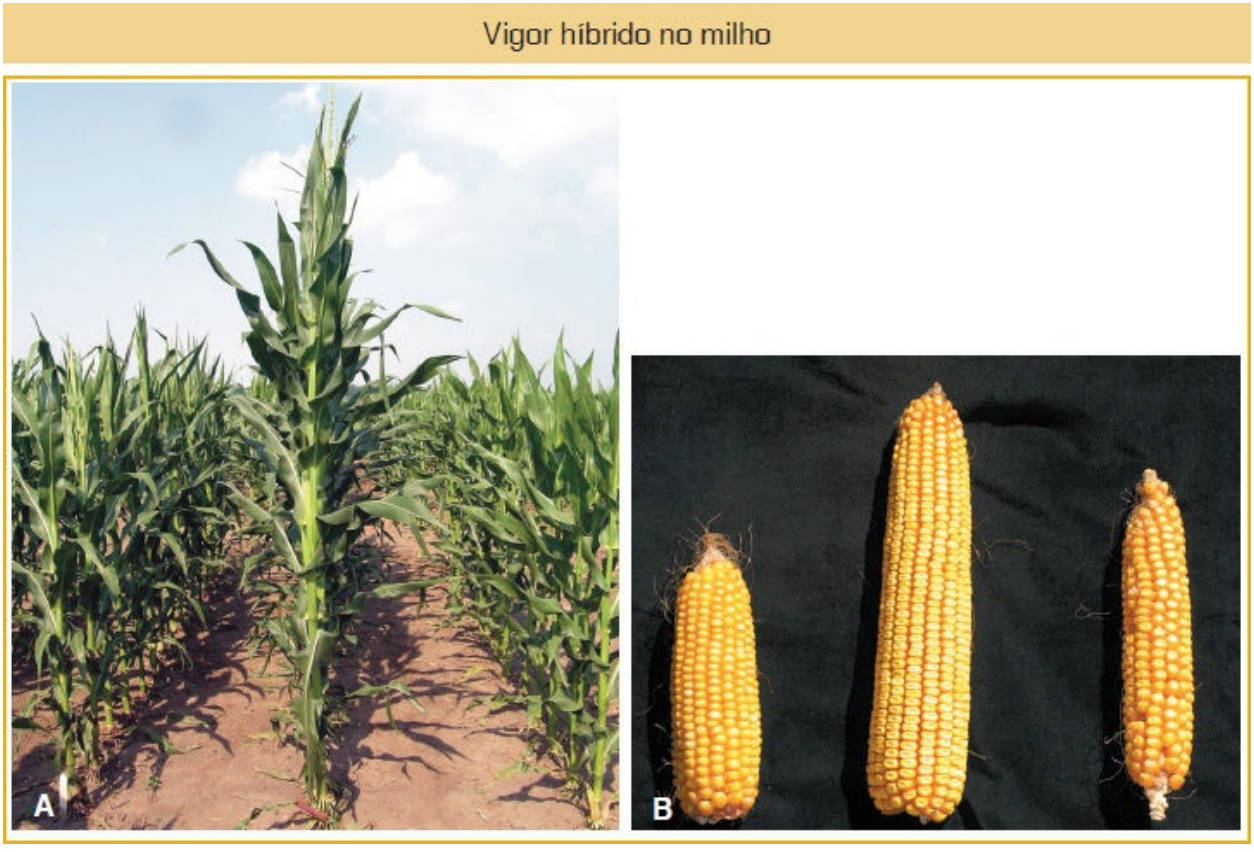


FIGURA 3.6 Híbridos heterozigotos múltiplos ladeados por duas linhagens puras cultivadas para a sua produção. A. As plantas. B. Espigas das mesmas plantas. (*A.* Foto cortesia de Jun Cao, Schnable Laboratory, Iowa State University; *B.* Deana Namuth-Covert, PhD, Univ. of Nebraska.)

A partir da perspectiva do usuário, existe outro aspecto negativo da utilização de híbridos. Após uma planta híbrida ter sido cultivada e produzida a sua safra para venda, não é realista manter algumas das sementes que ela produz e esperar que as suas sementes sejam igualmente vigorosas no próximo ano. O motivo é que, quando o híbrido é submetido à meiose, a distribuição independente dos diversos pares de genes misturados formará muitas combinações alélicas diferentes, e muito poucas dessas combinações serão aquelas do híbrido original. Por exemplo, o tetra-híbrido descrito anteriormente, quando autofecundado, produz 81 genótipos diferentes, dos quais apenas uma minoria será tetra-híbrida. Se presumirmos a distribuição independente, então, em relação a cada par de genes, a autofecundação produzirá metade de heterozigotos $A/a \rightarrow \frac{1}{4}\,A/A$, $\frac{1}{2}\,A/a$ e $\frac{1}{4}\,a/a$. Tendo em vista que existem quatro pares de genes nesse tetra-híbrido, a proporção da progênie que provavelmente será como o híbrido original A/a; B/b; C/c; D/d será de $(1/2)^4 = 1/16$.

> CONCEITO-CHAVE Alguns híbridos entre linhagens geneticamente diferentes demonstram vigor híbrido. Entretanto, a distribuição dos genes, quando o híbrido é submetido à meiose, rompe a combinação alélica favorável e, assim, poucos membros da próxima geração a apresentarão.

3.3 Base cromossômica da segregação independente

Assim como a segregação igual, a segregação independente de pares de genes em diferentes cromossomos é explicada pelo comportamento dos cromossomos durante a meiose. Considere um cromossomo que podemos denominar número 1; seus dois homólogos podem ser denominados 1′ e 1″. Se os cromossomos se alinham no equador, então 1′ pode se dirigir para o “norte” e 1″ para o “sul”, ou vice-versa. De modo semelhante, em relação a um cromossomo 2 com homólogos 2′ e 2″, 2′ pode se dirigir para o norte e 2″ para o sul, ou vice-versa. Portanto, o cromossomo 1′ pode acabar sendo empacotado com o cromossomo 2′ ou 2″, dependendo de quais cromossomos foram puxados na mesma direção.

A distribuição independente não é fácil de demonstrar por meio da observação da segregação cromossômica ao microscópio, tendo em vista que homólogos tais como 1′ e 1″ normalmente não aparentam ser diferentes, embora possam carregar uma pequena variação de sequência. Entretanto, a distribuição independente pode ser observada em determinados casos especializados. Um caso foi crucial no desenvolvimento histórico da teoria cromossômica.

Em 1913, Elinor Carothers encontrou uma situação cromossômica incomum em uma determinada espécie de gafanhoto — uma situação que possibilitou testar diretamente se os diferentes pares de cromossomos de fato segregam independentemente. Estudando as meioses nos testículos de gafanhotos, ela encontrou um gafanhoto no qual um "par" de cromossomos apresentava membros não idênticos. Tal par é denominado *heteromórfico*. Presumivelmente, os cromossomos demonstram homologia apenas parcial. Além disso, o mesmo gafanhoto apresentava outro cromossomo (não relacionado com o par heteromórfico) que absolutamente não apresentava um parceiro de pareamento. Carothers conseguiu utilizar esses cromossomos incomuns como marcadores citológicos visíveis do comportamento dos cromossomos durante a meiose. Ela triou visualmente muitas meioses e observou que havia dois padrões distintos, que estão demonstrados na Figura 3.7. Além disso, ela observou que os dois padrões eram igualmente frequentes. Em resumo, se mantivermos a segregação do par heteromórfico constante (em marrom na figura), então o cromossomo não pareado (roxo) pode se dirigir para qualquer polo com frequência igual, metade do tempo com a forma longa e metade do tempo com a forma curta. Em outras palavras, os conjuntos roxo e marrom segregaram-se de maneira independente. Embora esses obviamente não sejam cromossomos típicos, os resultados sugerem fortemente que diferentes cromossomos distribuem-se independentemente na primeira divisão da meiose.

Distribuição independente em organismos diploides

A base cromossômica da lei de segregação independente está formalmente diagramada na Figura 3.8, que ilustra como o comportamento separado de dois pares de cromossomos diferentes dá origem às proporções mendelianas de

1:1:1:1 dos tipos gaméticos esperados a partir da distribuição independente. A célula hipotética apresenta quatro cromossomos: um par de cromossomos homólogos longos (amarelo) e um par de cromossomos homólogos curtos (azul). O genótipo dos meiócitos é *A*/*a*; *B*/*b* e os dois pares alélicos, *A*/*a* e *B*/*b*, estão demonstrados em dois pares de cromossomos diferentes. As partes 4 e 4′ da Figura 3.8 demonstram a etapa-chave na distribuição independente: existem dois padrões de segregação alélica igualmente frequentes, um demonstrado em 4 e o outro em 4′. Em um caso, os alelos *A*/*A* e *B*/*B* são puxados em conjunto para dentro de uma célula, e os *a*/*a* e *b*/*b* são puxados para dentro da outra célula. No outro caso, os alelos *A*/*A* e *b*/*b* estão unidos na mesma célula e os alelos *a*/*a* e *B*/*B* também estão unidos na mesma célula. Os dois padrões resultam de duas ligações ao fuso igualmente frequentes dos centrômeros na anáfase I. Em seguida a meiose produz quatro células dos genótipos indicados de cada um desses padrões de segregação. Tendo em vista que os padrões de segregação 4 e 4′ são igualmente comuns, as células do produto meiótico dos genótipos *A*; *B*, *a*; *b*, *A*; *b* e *a*; *B* são produzidos em frequências iguais. Em outras palavras, a frequência de cada um dos quatro genótipos é de 1/4. Essa distribuição gamética é aquela postulada por Mendel em relação a um di-híbrido e é aquela que inserimos ao longo de uma borda do quadrado de Punnett na Figura 3.4. A fusão aleatória desses gametas resulta na proporção fenotípica da F_2 de 9:3:3:1.

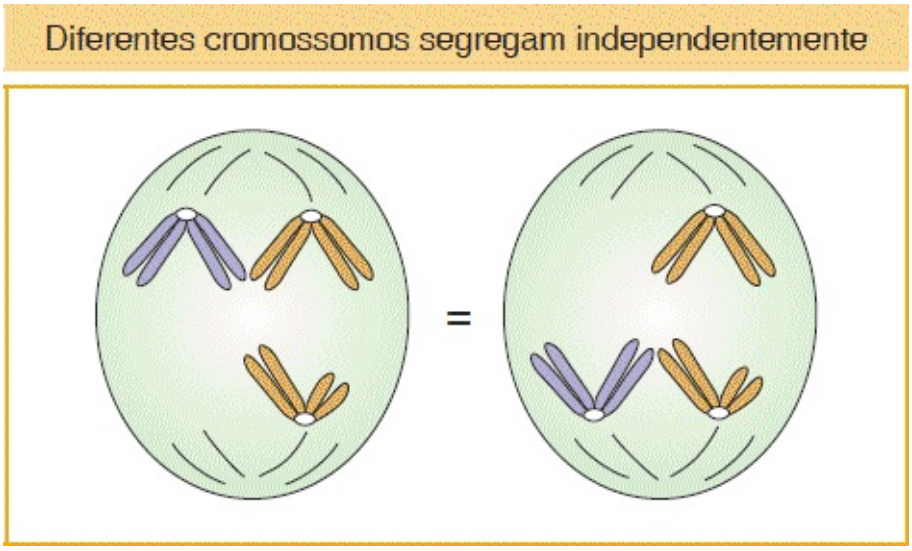


FIGURA 3.7 Carothers observou estes dois padrões igualmente frequentes, por meio dos quais um par heteromórfico (*marrom*) e um cromossomo não pareado (*roxo*) se movimentam para os gametas na meiose.

A distribuição independente dos cromossomos na meiose explica a proporção de Mendel

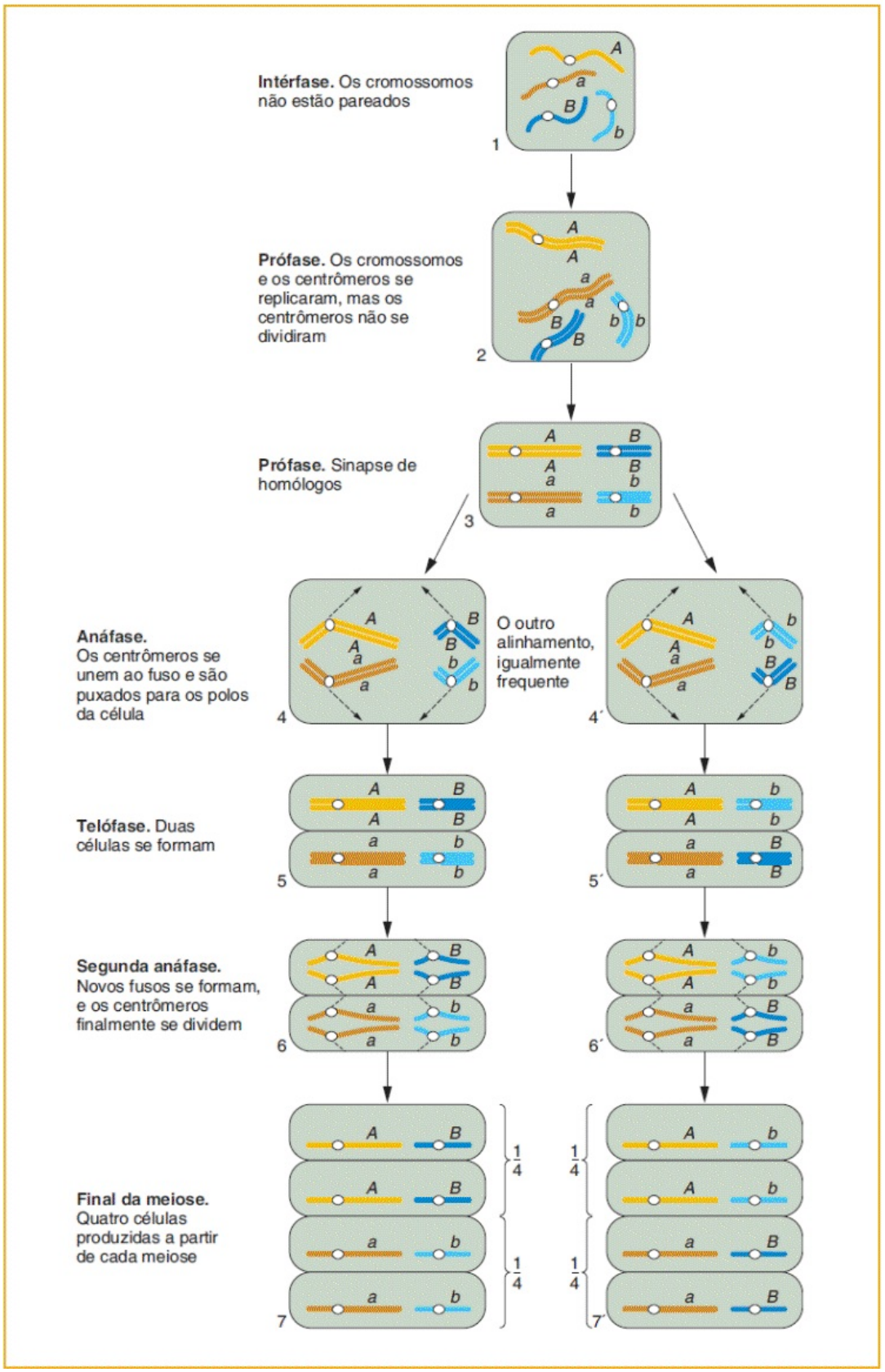


FIGURA 3.8 Meiose em uma célula diploide do genótipo *A*/*a*; *B*/*b*. O diagrama demonstra como a segregação e a distribuição dos diferentes pares de cromossomos dão origem à proporção gamética mendeliana de 1:1:1:1.

Segregação independente em organismos haploides

Nos fungos ascomicetos, de fato podemos inspecionar os produtos de um único meiócito para demonstrar diretamente a distribuição independente. Utilizemos o fungo filamentoso *Neurospora crassa* para ilustrar esse ponto (ver Organismo-modelo, adiante). Conforme observamos a partir dos exemplos anteriores de fungos, um cruzamento em *Neurospora* é realizado por meio da mistura de duas linhagens haploides parentais de tipo reprodutivo oposto. De modo similar àquele da levedura, o tipo reprodutivo é determinado por dois "alelos" de um gene — nessa espécie, denominados MAT-A e MAT-a. O modo por meio do qual um cruzamento é realizado está demonstrado na Figura 3.9.

Os produtos da meiose nos fungos são esporos sexuados. Relembre que os *ascomicetos* (que incluem *Neurospora* e *Saccharomyces*) são únicos no sentido em que, em relação a qualquer determinado meiócito, os esporos são mantidos juntos em um saco membranoso denominado *asco*. Portanto, em relação a esses organismos, os produtos de uma única meiose podem ser recuperados e testados. No bolor laranja do pão *Neurospora*, os fusos nucleares das meioses I e II não se sobrepõem dentro do asco com formato de charuto e, assim, os quatro produtos de um único meiócito se encontram em uma linha reta (Figura 3.10 A. Além disso, em virtude de algum motivo não compreendido, existe uma *mitose pós-meiose*, que também não demonstra sobreposição do fuso. Portanto, a meiose e a mitose extra resultam em um asco linear que contém 8 *ascósporos*, ou uma *óctade*. Em um meiócito heterozigoto *A*/*a*, se não houver *crossings* entre os genes e seu centrômero (ver Capítulo 4), haverá então dois blocos adjacentes de ascósporos, quatro de *A* e quatro de *a* (Figura 3.10 B).

Agora podemos examinar um di-híbrido. Façamos um cruzamento entre dois mutantes distintos que apresentam mutações em diferentes genes em diferentes cromossomos. Ao presumir que os *loci* dos genes mutados estão muito próximos

de seus respectivos centrômeros, evitamos complicações em virtude do *crossing over* entre os *loci* e os centrômeros. O primeiro mutante é albino (*a*), que contrasta com o tipo selvagem normal rosa (a^+). O segundo mutante é biscoito (*b*), que apresenta uma colônia muito compacta com formato semelhante a um biscoito, que contrasta com a colônia plana e espalhada do tipo selvagem (b^+). Presumiremos que os dois mutantes são de tipo reprodutivo oposto. Portanto, o cruzamento é:

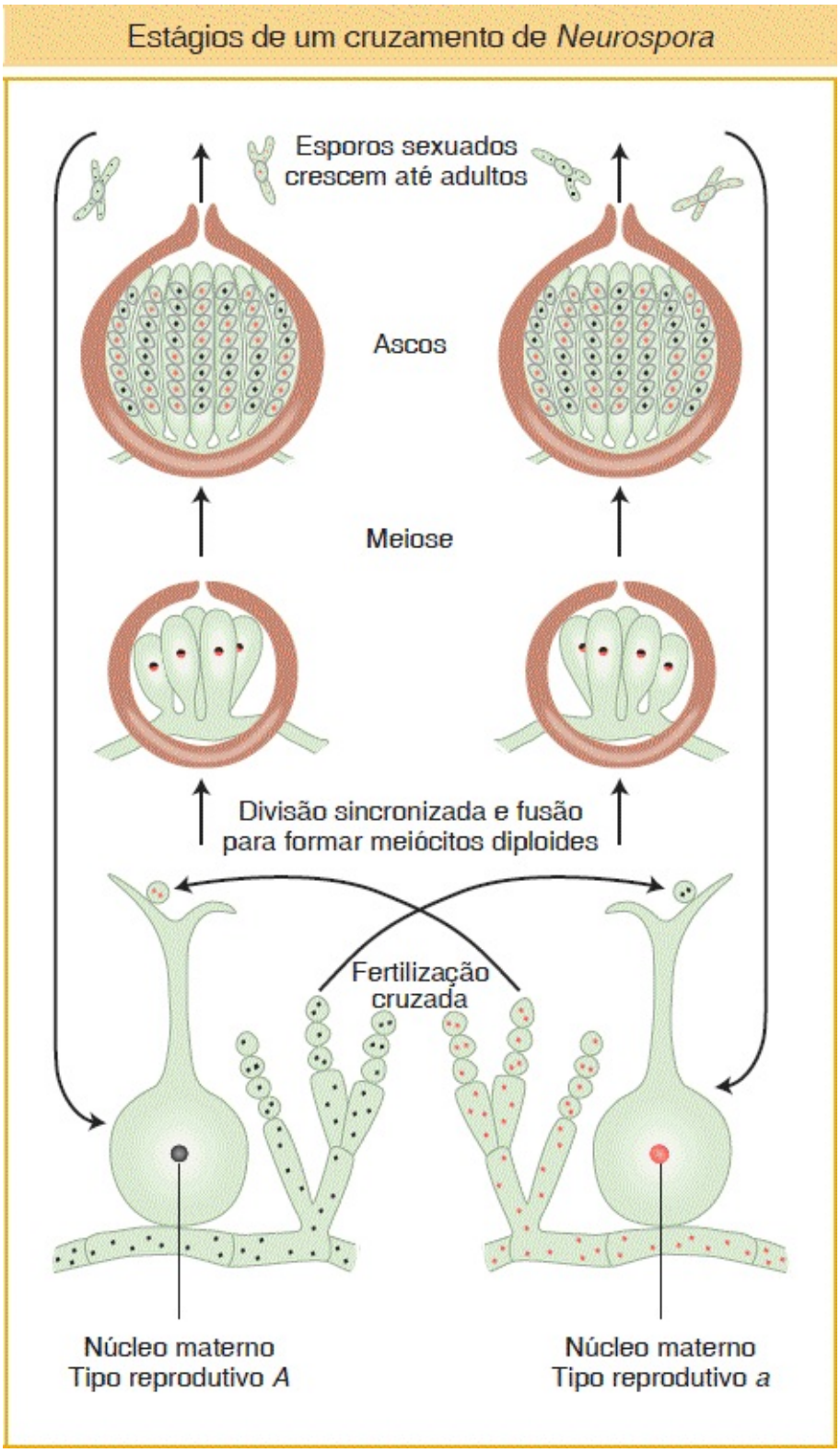


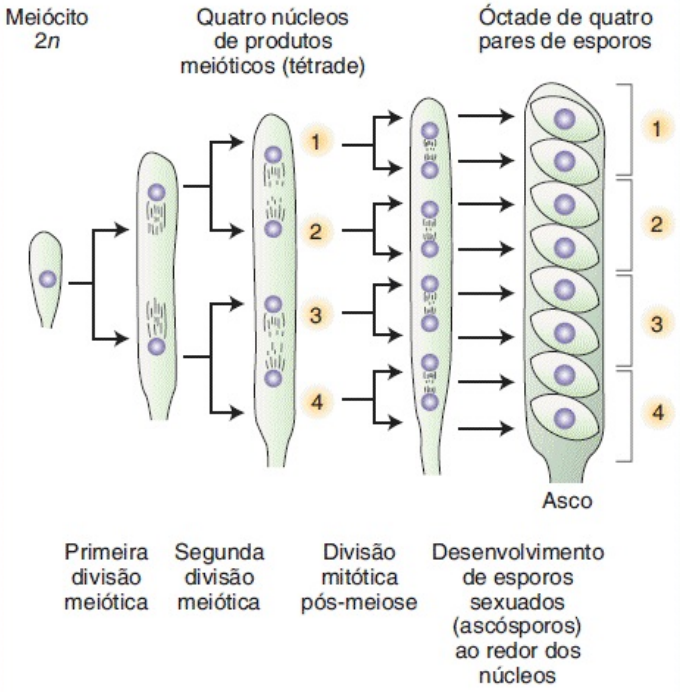
FIGURA 3.9 O ciclo de vida de *Neurospora crassa*, o bolor laranja do pão. A autofertilização não é possível nesta espécie: existem dois tipos reprodutivos, determinados pelos alelos *A* e *a* de um gene, e qualquer um

deles pode atuar como “fêmea”. Um esporo assexuado de tipo reprodutivo oposto se funde com um filamento receptivo, e um núcleo do esporo assexuado se difunde através do filamento até parear-se com um núcleo feminino nas células. O par *A* e *a* em seguida é submetido a mitoses sincronizadas, finalmente se fundindo para formar meiócitos diploides.

$$a; b^{+} \times a^{+}; b$$

A meiose linear de *Neurospora*

A. Divisões nucleares



B. Segregação alélica

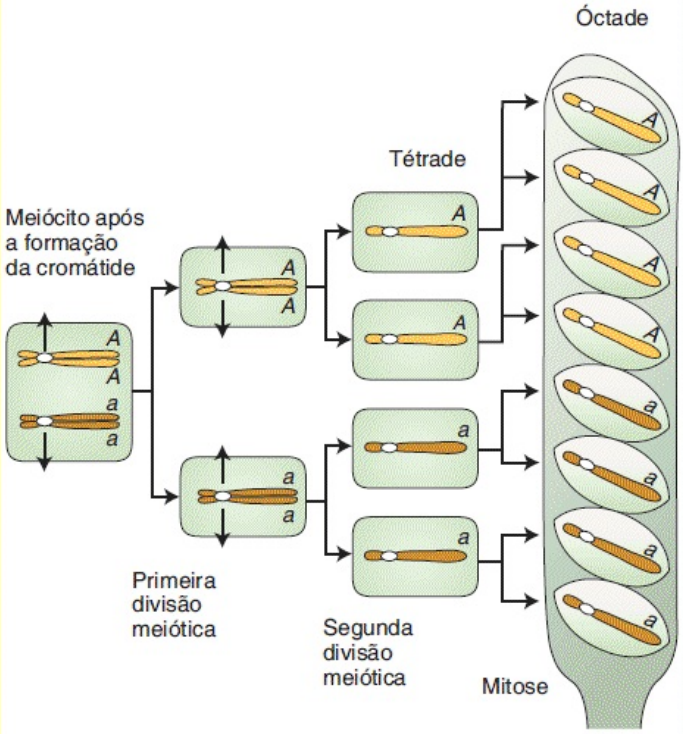


FIGURA 3.10 *Neurospora* é um sistema-modelo ideal para o estudo da segregação alélica na meiose. A. Os quatro produtos da meiose (tétrade) são submetidos à mitose para produzir uma óctade. Os produtos estão contidos dentro de um asco. B. Um meiócito *A*/*a* é submetido à meiose, seguida pela mitose, resultando em números iguais de produtos *A* e *a* e demonstrando o princípio da segregação igual.

Em virtude da ligação aleatória ao fuso, os 8 tipos de óctades a seguir serão igualmente frequentes:

a^+; b	a; b
a^+; b	a; b
a^+; b	a; b
a^+; b	a; b
a; b^+	a^+; b^+
a; b^+	a^+; b^+
a; b^+	a^+; b^+
a; b^+	a^+; b^+
50%	**50%**

A frequência igual desses dois tipos é uma demonstração convincente da distribuição independente que ocorre em meiócitos individuais.

Segregação independente de combinações de genes autossômicos e ligados ao X

O princípio da segregação independente é algo útil na análise dos genótipos que são heterozigotos em relação a ambos os genes autossômicos e ligados ao X. Os autossomos e os cromossomos sexuais são movimentados independentemente por fibras do fuso ligadas aleatoriamente aos seus centrômeros, justamente como com dois pares de autossomos diferentes. São produzidas algumas proporções de di-

híbridos interessantes. Vejamos um exemplo da *Drosophila*. O cruzamento é entre uma fêmea com asas vestigiais (autossômico recessivo, *vg*) e um macho com olhos brancos (recessivo ligado ao X, *w*). Simbolicamente, o cruzamento é:

$$\text{♀ } vg/vg;\ +/+ \times \text{♂ } +/+;\ w/\text{Y}$$

A F_1 será:

Fêmeas de genótipo	$+/vg;\ +/w$
Machos de genótipo	$+/vg;\ +/\text{Y}$

Essas moscas da F_1 têm de ser intercruzadas para a obtenção de uma F_2. Tendo em vista que o cruzamento é um cruzamento mono-híbrido em relação ao gene autossômico vestigial, ambos os sexos da F_2 demonstrarão:

Fêmeas e machos	$\frac{3}{4}$ +/— (tipo selvagem)
	$\frac{1}{4}$ *vg*/*vg* (vestigial)

Em relação ao gene dos olhos brancos ligado ao X, as proporções serão como segue:

Fêmeas	$\frac{1}{2}$ +/+ e $\frac{1}{2}$ +/*w* (todas do tipo selvagem)
Machos	$\frac{1}{2}$ +/Y (tipo selvagem) e $\frac{1}{2}$ *w*/Y (brancos)

Se os genes autossômicos e ligados ao X forem combinados, as proporções fenotípicas da F_2 serão:

Fêmeas	$\frac{3}{4}$ totalmente do tipo selvagem
	$\frac{1}{4}$ vestigiais
Machos	$\frac{3}{8}$ totalmente do tipo selvagem ($\frac{3}{4} \times \frac{1}{2}$)
	$\frac{3}{8}$ brancos ($\frac{3}{4} \times \frac{1}{2}$)
	$\frac{1}{8}$ vestigiais ($\frac{1}{4} \times \frac{1}{2}$)
	$\frac{1}{8}$ vestigiais, brancos ($\frac{1}{4} \times \frac{1}{2}$)

Portanto, observamos uma proporção da progênie que revela elementos claros de ambas as heranças, autossômica e ligada ao X.

Recombinação

A distribuição independente dos genes na meiose é um dos principais modos por meio dos quais um organismo produz novas combinações de alelos. A produção de novas combinações de alelos é formalmente denominada **recombinação**.

Existe uma concordância geral de que a vantagem evolutiva da produção de novas combinações de alelos é que ela produz variação como a matéria-prima para a seleção natural. A recombinação é um princípio crucial em genética, em parte em virtude da sua relevância para a evolução, mas também em virtude de sua utilização na análise genética. Ela é particularmente útil para a análise dos padrões de herança de genótipos multigênicos. Nesta seção, definiremos a recombinação de tal modo que possamos reconhecê-la em resultados experimentais, e mostramos o modo como a recombinação é analisada e interpretada.

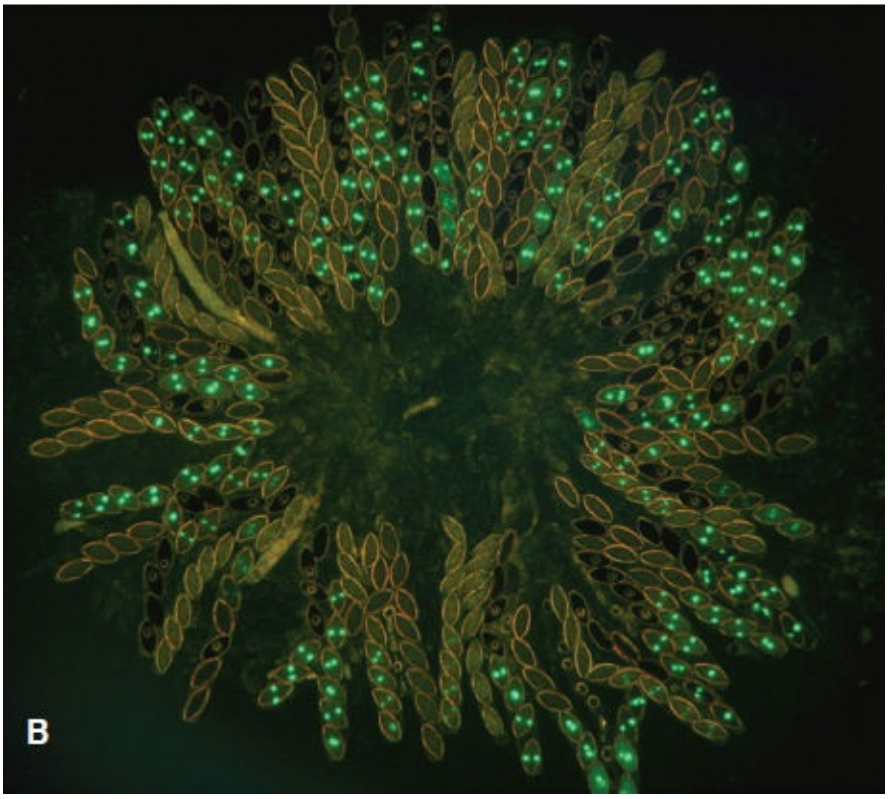
ORGANISMO-MODELO *Neurospora*

Neurospora crassa foi um dos primeiros microrganismos eucarióticos a ser adotado pelos geneticistas como um organismo-modelo. É um fungo haploide ($n = 7$), que pode ser observado crescendo na vegetação morta em muitas partes do mundo. Quando um esporo assexuado (haploide) germina, ele produz uma estrutura tubular que se estende rapidamente por meio do crescimento das extremidades e que origina diversos ramos laterais. O resultado é uma massa de filamentos ramificados (denominados *hifas*), que constituem uma colônia. As hifas não apresentam paredes cruzadas, de modo que uma colônia é essencialmente uma célula que contém muitos núcleos haploides. De uma colônia brotam milhões de esporos assexuados, que conseguem se dispersar e repetir o ciclo assexuado.

As colônias assexuadas são mantidas facilmente e de modo não dispendioso em laboratório ou em meio definido de sais inorgânicos mais uma fonte de

energia, como açúcar. (Um gel inerte, tal como ágar, é adicionado para proporcionar uma superfície firme.) O fato de que *Neurospora* consegue sintetizar quimicamente todas as suas moléculas essenciais a partir de um referido meio simples levou os geneticistas bioquímicos (com início com George Beadle e Edward Tatum; ver Capítulo 6) a escolhê-lo para os estudos das vias sintéticas. Os geneticistas elaboraram as etapas nessas vias por meio da introdução de mutações e da observação de seus efeitos. O estado haploide de *Neurospora* é ideal para a referida análise mutacional, tendo em vista que os alelos mutantes são sempre expressos diretamente no fenótipo.

Neurospora apresenta dois tipos reprodutivos, MAT-A e MAT-*a*, que podem ser considerados como "sexos" simples. Quando colônias de tipos reprodutivos diferentes entram em contato, suas paredes celulares e seus núcleos se fundem, resultando em muitos núcleos diploides temporários, cada um dos quais sofre meiose. Os quatro produtos haploides de uma meiose permanecem juntos em um saco denominado *asco*. Cada um desses produtos da meiose é submetido a uma divisão mitótica adicional, resultando em oito ascósporos dentro de cada asco. Os ascósporos germinam e produzem colônias exatamente iguais àquelas produzidas pelos esporos assexuados. Portanto, os referidos fungos *ascomicetos* são ideais para o estudo da segregação e da recombinação dos genes em meioses individuais.



O fungo ***Neurospora crassa.*** **A.** Colônias laranja de *Neurospora* crescendo na cana-de-açúcar. Na natureza, as colônias de *Neurospora* são observadas com mais frequência após incêndios, que ativam os ascósporos dormentes. (Campos de cana-de-açúcar são queimados para remover a folhagem antes da colheita dos caules da cana.) **B.** Óctades de *Neurospora* em desenvolvimento a partir de um cruzamento do tipo selvagem com uma linhagem que carreia um

alelo modificado da proteína fluorescente verde de água-viva fundido à histona. As óctades demonstram a segregação mendeliana esperada de 4:4 da fluorescência. Em alguns esporos, o núcleo foi dividido mitoticamente para formar dois; finalmente, cada esporo conterá diversos núcleos. (*A.* David J. Jacobson, Ph.D.; *B.* Namboori B. Raju, Stanford University.)

A recombinação é observada em uma diversidade de situações biológicas, mas por enquanto a definimos em relação à meiose.

> **Recombinação meiótica** é qualquer processo meiótico que gera um produto haploide com novas combinações dos alelos carreados pelos genótipos haploides que se uniram para formar o meiócito.

A definição aparentemente prolixa é realmente bastante simples; ela demonstra o importante ponto de que detectamos a recombinação por meio da comparação das *entradas* na meiose com as *saídas* (Figura 3.11). As entradas são os dois genótipos haploides que se combinam para formar o meiócito, a célula diploide que sofre meiose. Em relação aos seres humanos, as entradas são o ovócito e o espermatozoide parentais. Eles se unem para formar um zigoto diploide, que se divide para produzir todas as células corporais, incluindo os meiócitos que são armazenados dentro das gônadas. Os genótipos de saída são os produtos haploides da meiose. Em seres humanos, esses produtos haploides são os ovócitos ou os espermatozoides da própria pessoa. Qualquer produto meiótico que apresente uma nova combinação dos alelos fornecidos por meio dos dois genótipos da entrada por definição é um **recombinante**.

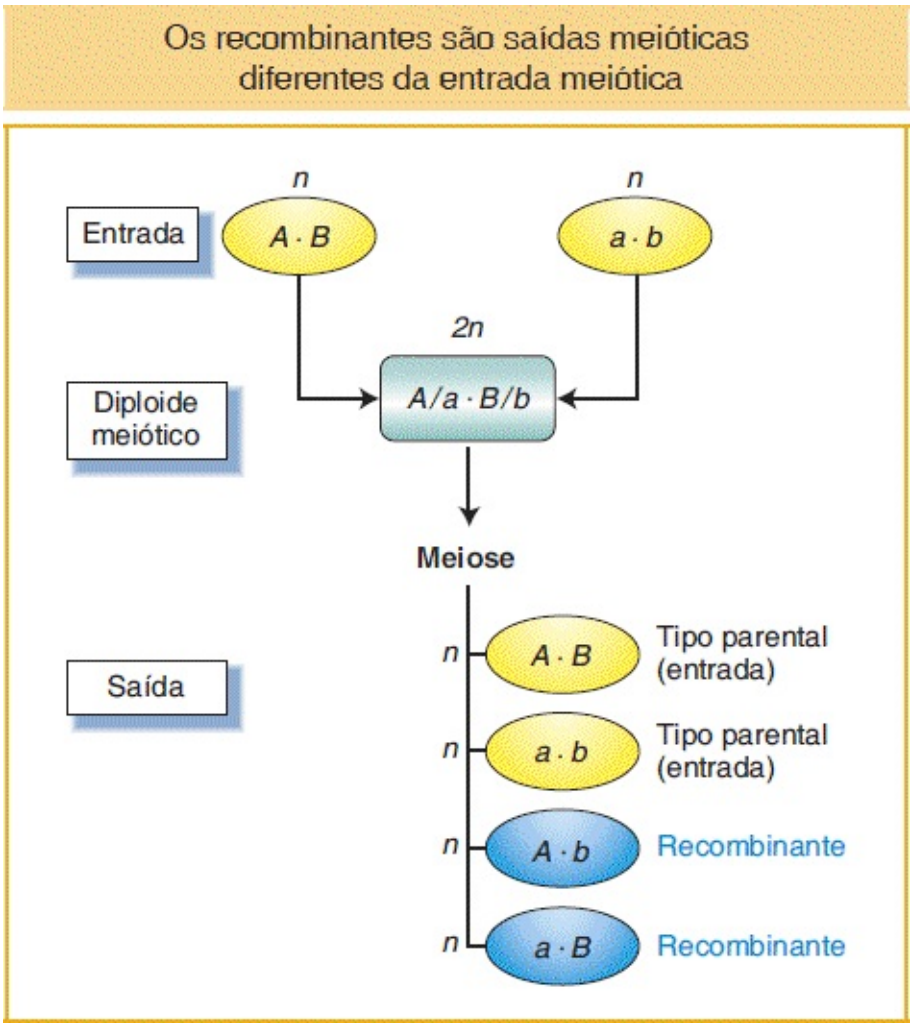


FIGURA 3.11 Os recombinantes (*azuis*) são aqueles produtos da meiose com combinações alélicas diferentes daquelas das células haploides que formaram o diploide meiótico (*amarelas*). Observe que os genes *A*/*a* e *B*/*b* estão demonstrados em separado por um ponto, tendo em vista que podem estar no mesmo cromossomo ou em cromossomos diferentes.

> CONCEITO-CHAVE A meiose gera recombinantes, que são produtos meióticos haploides com novas combinações dos alelos carreados pelos genótipos haploides que se uniram para formar o meiócito.

Primeiramente, vejamos como os recombinantes são detectados experimentalmente. A detecção dos recombinantes em organismos com ciclos de vida haploides, tais como fungos ou algas, é direta. Os tipos de entrada e saída

em ciclos de vida haploides são os genótipos dos indivíduos, em vez dos gametas e, portanto, podem ser inferidos diretamente a partir dos fenótipos. A Figura 3.11 pode ser considerada como um resumo da simples detecção de recombinantes em organismos com ciclos de vida haploides. A detecção de recombinantes em organismos com ciclos de vida diploides é mais complicada. Os tipos de entrada e saída em ciclos diploides são os gametas. Portanto, precisamos conhecer os genótipos de ambos os gametas de entrada e saída para detectar os recombinantes com um ciclo diploide. Embora não possamos detectar os genótipos dos gametas de entrada ou saída diretamente, podemos inferir esses genótipos por meio da utilização das técnicas adequadas:

- *Para conhecer os gametas de entrada*, utilizamos genitores diploides puros, tendo em vista que eles podem produzir apenas um tipo gamético
- *Para detectar os gametas de saída recombinantes*, realizamos o cruzamento-teste do indivíduo diploide e observamos a sua progênie (Figura 3.12).

Uma descendência de um cruzamento-teste que surge de um produto recombinante da meiose também é denominada *recombinante*. Observe, novamente, que o cruzamento-teste possibilita que nos concentremos em *uma* meiose e evita a ambiguidade. A partir da *autofecundação* da F_1 na Figura 3.12, por exemplo, uma descendência recombinante $A/A \cdot B/b$ não poderia ser diferenciada de $A/A \cdot B/B$ sem cruzamentos adicionais.

Uma parte central da análise da recombinação é a frequência de recombinantes. Um motivo para enfocar na frequência de recombinantes é que seu valor numérico é um teste conveniente para saber se dois genes estão em cromossomos diferentes. Os recombinantes são produzidos por meio de dois processos celulares diferentes: a distribuição independente dos genes em cromossomos diferentes (neste capítulo) e o *crossing over* entre genes no mesmo cromossomo (ver Capítulo 4). A *proporção* de recombinantes é uma ideia-chave aqui, tendo em vista que o valor diagnóstico pode nos informar se os genes estão em cromossomos diferentes. Aqui, lidaremos com a distribuição independente.

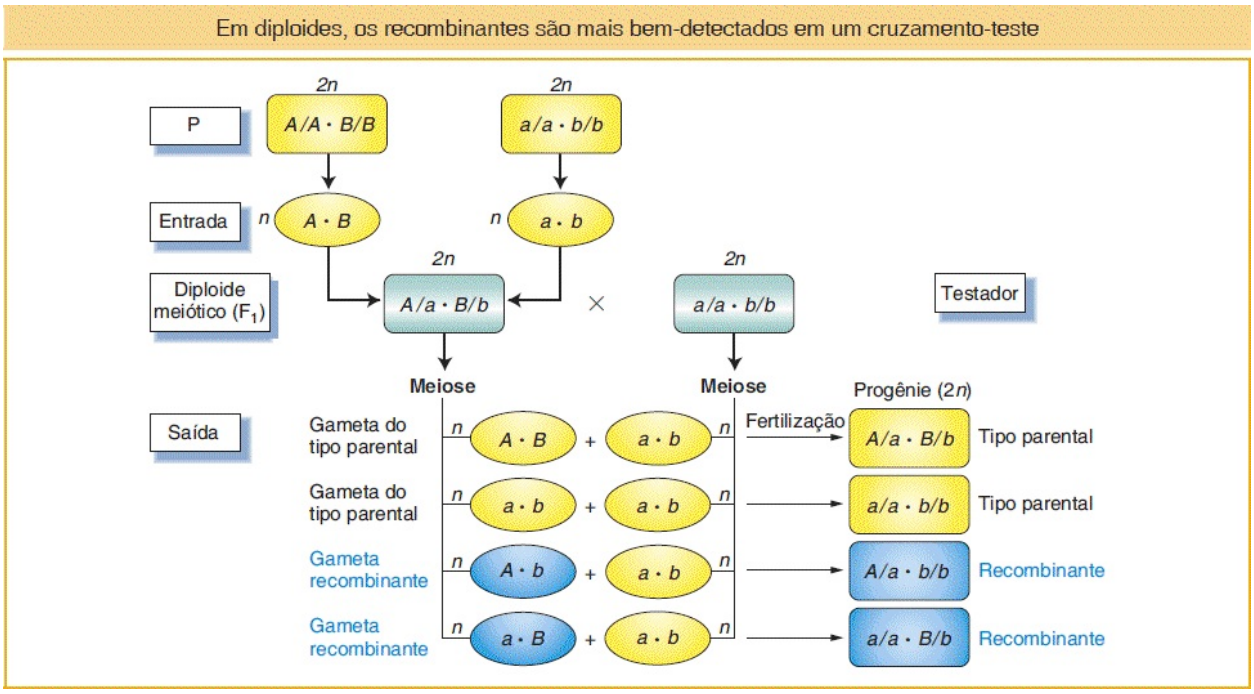


FIGURA 3.12 Os produtos recombinantes de uma meiose diploide são na maior parte prontamente detectados em um cruzamento de um heterozigoto e um testador recessivo. Observe que a Figura 3.11 é repetida como parte deste diagrama.

Em relação aos genes em cromossomos separados, os recombinantes são produzidos por meio da distribuição independente, conforme demonstrado na Figura 3.13. Novamente, observamos a proporção de 1:1:1:1 que observamos anteriormente, mas agora a progênie do cruzamento-teste é classificada como recombinante ou semelhante aos tipos de entrada P (parental). Estabelecida desse modo, a proporção de recombinantes é claramente $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, ou 50% da progênie total. Portanto, observamos que a distribuição independente na meiose produz uma frequência de recombinantes de 50%. Se observarmos uma frequência de recombinantes de 50% em um cruzamento-teste, podemos inferir que os dois genes em estudo se distribuem independentemente. A interpretação mais simples e mais provável da distribuição independente é que os dois genes estão em pares de cromossomos separados. (Entretanto, devemos observar que genes que estão muito distantes no *mesmo* par de cromossomos podem virtualmente ser distribuídos independentemente e produzir o mesmo resultado; ver Capítulo 4.)

> CONCEITO-CHAVE Uma frequência de recombinantes de 50% indica que os genes são distribuídos independentemente e que mais provavelmente estão em cromossomos diferentes.

3.4 Herança poligênica

Até agora, a nossa análise neste livro enfocou diferenças monogênicas, com a utilização de fenótipos nitidamente contrastantes, tais como pétalas vermelhas *versus* brancas, sementes lisas *versus* rugosas, e *Drosophila* com asas longas *versus* vestigiais. Entretanto, uma grande parte da variação observada na natureza é *contínua*, na qual um fenótipo pode assumir qualquer valor mensurável entre dois extremos. Estatura, peso e intensidade da cor são exemplos dos referidos *fenótipos métricos*, ou *quantitativos*. Em geral, quando o valor métrico é plotado contra a frequência em uma população natural, a curva de distribuição apresenta um formato semelhante a um sino (Figura 3.14). O formato de sino ocorre em virtude de os valores médios na parte intermediária serem os mais comuns, enquanto os valores extremos são raros. Primeiramente, é difícil observar como as distribuições contínuas podem ser influenciadas por genes herdados de modo mendeliano. Afinal, toda análise mendeliana é facilitada por meio da utilização de categorias claramente distinguíveis. Entretanto, veremos nesta seção que a distribuição independente de diversos a muitos genes heterozigotos que afetam um traço contínuo pode produzir uma curva em sino.

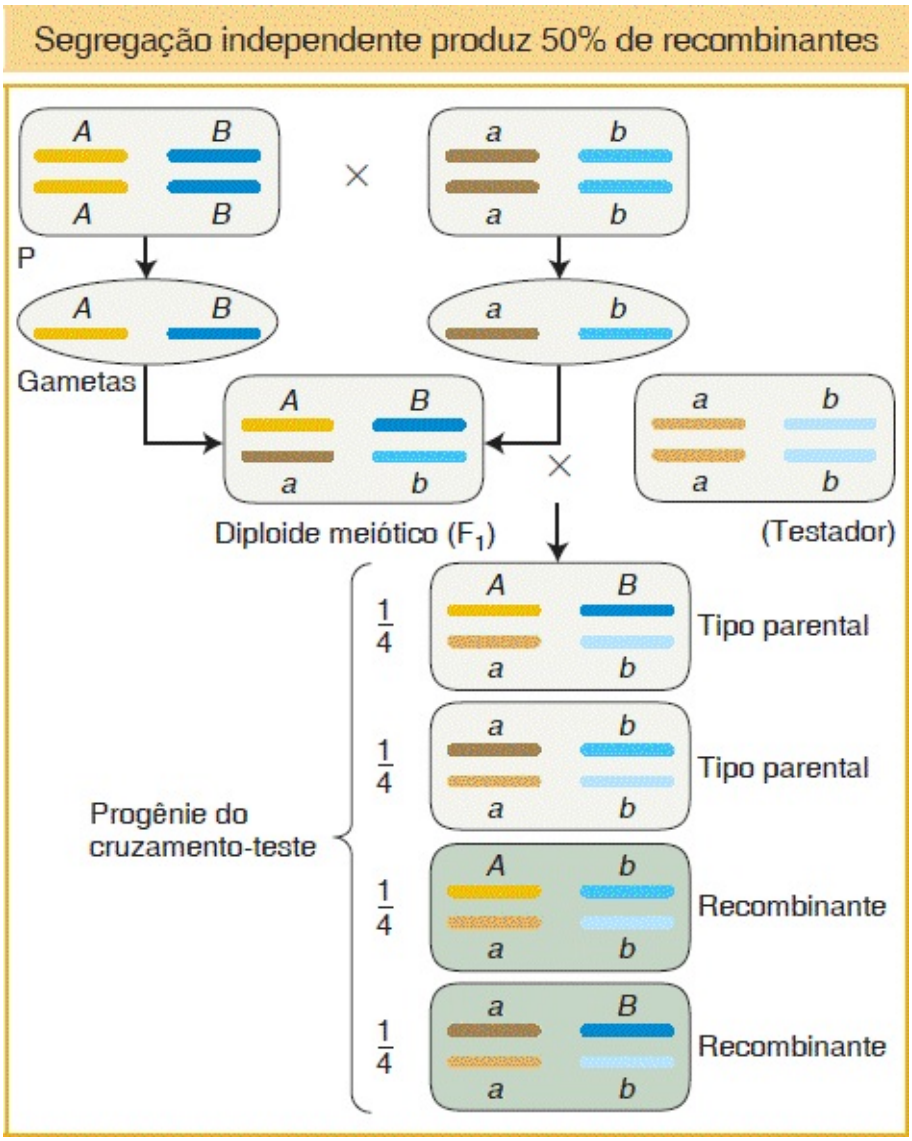


FIGURA 3.13 Este diagrama demonstra dois pares de cromossomos de um organismo diploide com *A* e *a* em um par e *B* e *b* no outro. A segregação independente produz uma frequência de recombinantes de 50%. Observe que poderíamos representar a situação haploide por meio da remoção das gerações parentais (P) e do testador.

É claro que muitos casos de variação contínua apresentam uma base puramente ambiental, pouco afetada pela genética. Por exemplo, uma população de plantas geneticamente homozigotas cultivadas em uma área de terreno com frequência demonstra uma curva sinusoidal em relação à altura, com as plantas menores ao redor das extremidades do gráfico e as plantas maiores na parte intermediária. Essa variação pode ser explicada apenas por fatores ambientais, tais como

umidade e quantidade de fertilizante aplicada. Entretanto, muitos casos de variação contínua apresentam uma base genética. A cor da pele humana é um exemplo: todos os graus de cor escura da pele podem ser observados em populações de diferentes partes do mundo, e essa variação claramente apresenta um componente genético. Nos referidos casos, diversos a muitos alelos interagem com um efeito mais ou menos aditivo. Os genes que interagem subjacentes à variação contínua hereditária são denominados **poligenes** ou ***loci* de traço quantitativo (QTL)**. (O termo *locus de traço quantitativo* requer uma definição: *quantitativo* é mais ou menos sinônimo de contínuo; *traço* é mais ou menos sinônimo de característica ou propriedade; *locus*, que literalmente significa local em um cromossomo, é mais ou menos sinônimo de gene.) Os poligenes, ou QTL em relação ao mesmo traço, estão distribuídos por todo o genoma; em muitos casos, estão em cromossomos diferentes e demonstram distribuição independente.

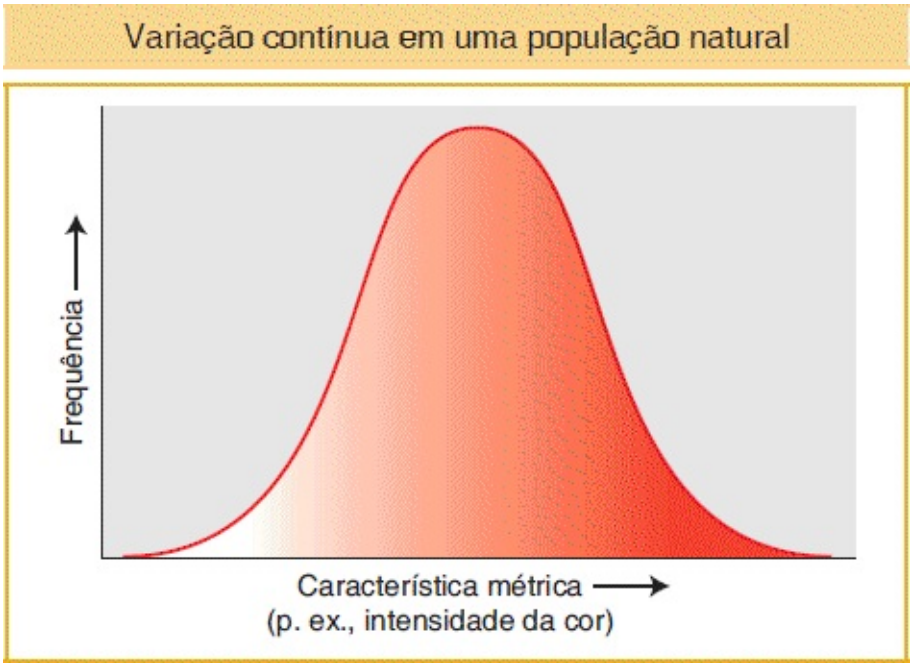


FIGURA 3.14 Em uma população, um caractere métrico, tal como a intensidade da cor, pode assumir muitos valores. Portanto, a distribuição ocorre na forma de uma curva uniforme, com os valores mais comuns representando o ponto alto da curva. Se a curva for simétrica, ela tem o formato de um sino, conforme demonstrado.

Vejamos como a herança de diferentes poligenes heterozigotos (até mesmo

dois) pode gerar uma curva de distribuição sinusoidal. Podemos considerar um modelo simples que foi utilizado originalmente para explicar a variação contínua no grau de vermelhidão em sementes de trigo. O trabalho foi realizado por Hermann Nilsson-Ehle no início do século 20. Presumiremos dois pares de genes que se distribuem independentemente, R_1/r_1 e R_2/r_2. Ambos, R_1 e R_2, contribuem para a vermelhidão da semente do trigo. Cada "dose" de um alelo *R* em cada gene é aditiva, o que significa que ela aumenta o grau de vermelhidão de modo proporcional. Um cruzamento ilustrativo é a autofecundação de um di-híbrido R_1/r_1; R_2/r_2. Ambos os *gametas* masculino e feminino demonstrarão proporções genotípicas como segue:

R_1; R_2	2 doses de vermelho
R_1; r_2	1 dose de vermelho
r_1; R_2	1 dose de vermelho
r_1; r_2	0 dose de vermelho

Em geral, nessa população de gametas, um quarto apresenta duas doses, metade apresenta uma dose, e um quarto apresenta zero dose. A união de ambos os gametas masculino e feminino que demonstra esse arranjo de doses de *R* está ilustrada na Figura 3.15. A quantidade de doses na progênie varia de quatro (R_1/R_1; R_2/R_2) a zero (r_1/r_1; r_2/r_2), com todos os valores entre elas.

As proporções na grade da Figura 3.15 podem ser desenhadas como um histograma, conforme demonstrado na Figura 3.16. O formato do histograma pode ser considerado como um arcabouço, que pode ser o fundamento de base para a curva de distribuição sinusoidal. Quando essa análise da vermelhidão nas sementes de trigo foi originalmente realizada, foi observada uma variação dentro das classes que alegadamente representavam um nível de "dose" de poligene. Presumivelmente, essa variação dentro de uma classe é o resultado de diferenças ambientais. Portanto, pode-se considerar que o ambiente contribui de um modo que arredonda as extremidades afiladas do histograma de barras, resultando em

uma curva suave com formato de sino (a linha vermelha no histograma). Se o número de poligenes for aumentado, o histograma aproxima-se mais de uma suave distribuição contínua. Por exemplo, em relação a uma característica determinada por três poligenes, o histograma é conforme demonstrado na Figura 3.17.

Em nossa ilustração, utilizamos um autocruzamento di-híbrido para demonstrar como o histograma é produzido. Mas como o nosso exemplo é relevante para o que está ocorrendo em populações naturais? Afinal, nem todos os cruzamentos podem ser desse tipo. Não obstante, se os alelos em cada par de genes forem aproximadamente iguais em frequência na população (p. ex., R_1 é quase tão comum quanto r_1), então pode ser dito que o cruzamento di-híbrido representa um cruzamento médio em relação a uma população na qual dois poligenes estão segregando.

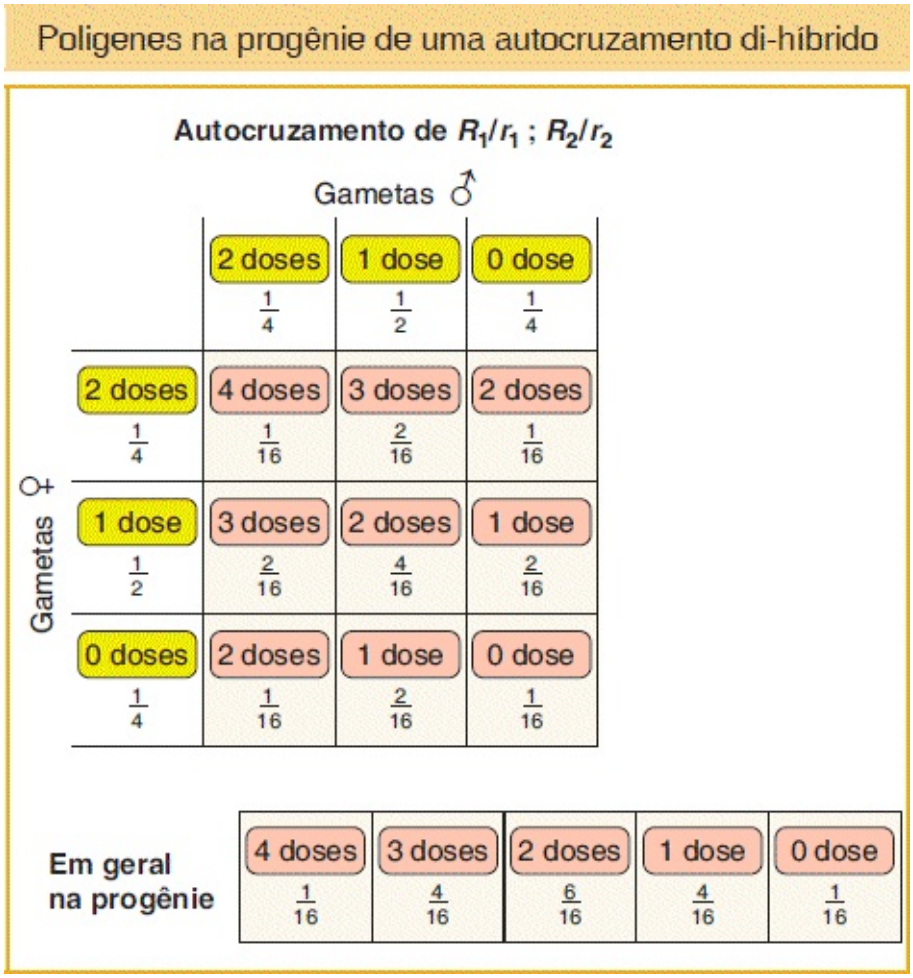


FIGURA 3.15 A progênie de um autocruzamento di-híbrido em relação a dois poligenes pode ser expressa como quantidades de “doses” alélicas aditivas.

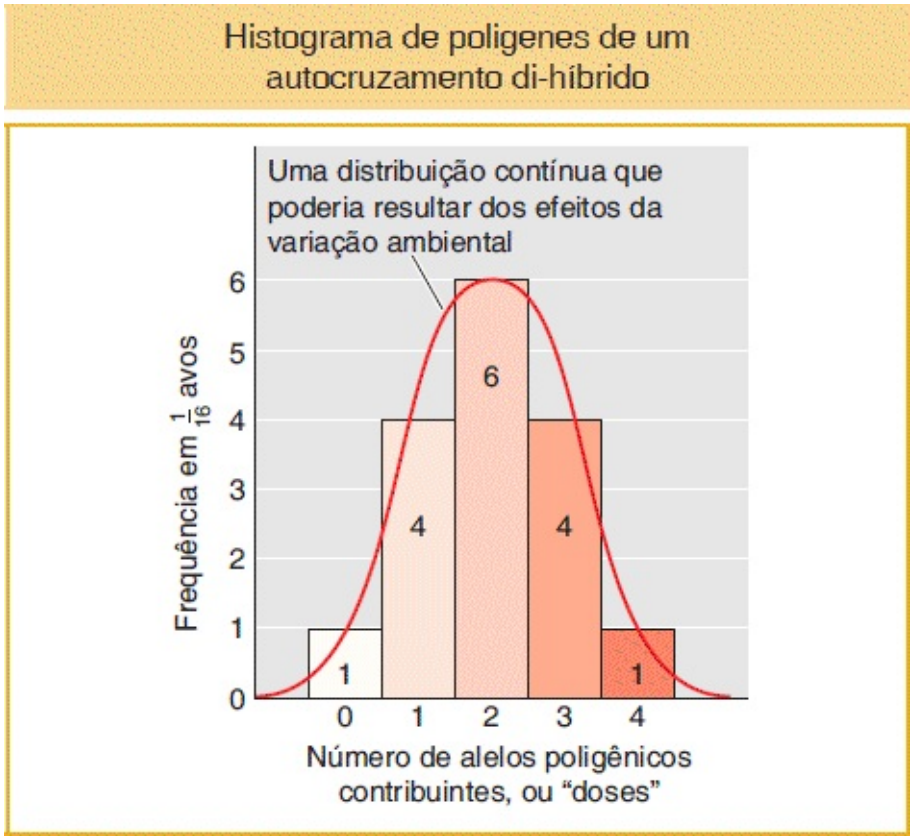


FIGURA 3.16 A progênie demonstrada na Figura 3.15 pode ser representada como um histograma de frequência de alelos poligênicos contribuintes (“doses”).

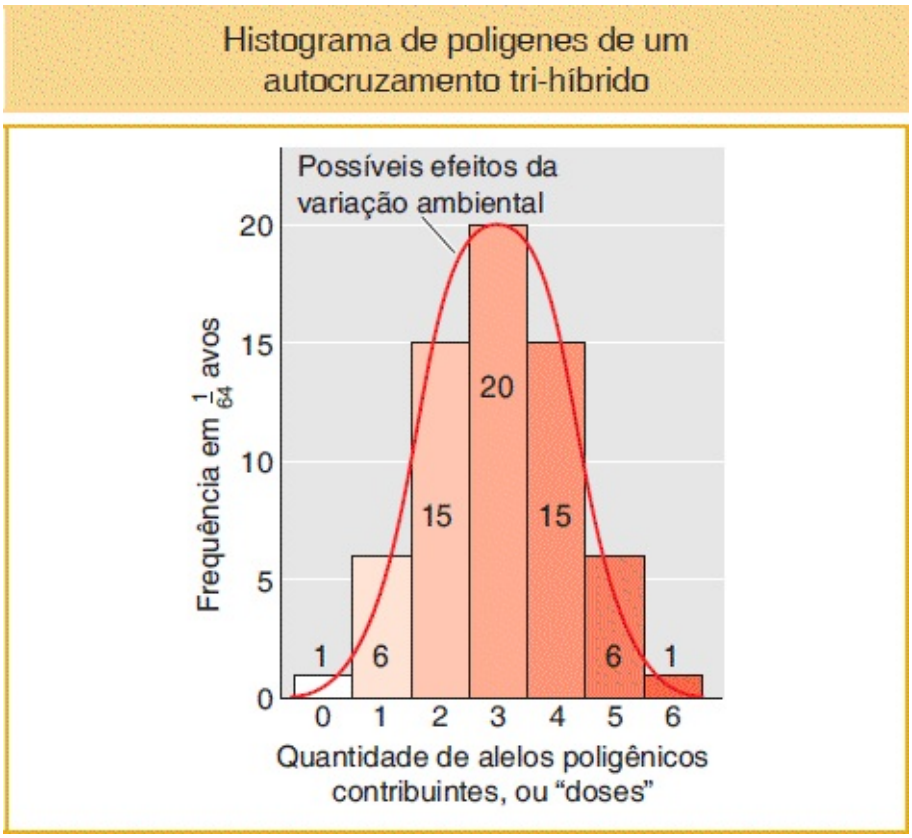


FIGURA 3.17 A progênie de um tri-híbrido poligênico pode ser representada por um gráfico de histograma de frequência de alelos poligênicos contribuintes ("doses").

A identificação de poligenes e a compreensão sobre como eles atuam e interagem são desafios importantes para os geneticistas no século 21. A identificação de poligenes será especialmente importante na medicina. Acredita-se que muitas doenças humanas comuns, tais como a aterosclerose e a hipertensão, apresentem um componente poligênico. Nesse caso, uma total compreensão sobre essas condições, que afetam grandes proporções de populações humanas, requer uma compreensão sobre esses poligenes, sua herança e função. Atualmente, diversas abordagens moleculares podem ser aplicadas à tarefa de encontrar poligenes. Consideraremos algumas nos capítulos subsequentes. Observe que os poligenes não são considerados uma classe funcional especial de genes. Eles são identificados como um grupo apenas no sentido em que apresentam alelos que contribuem para a variação contínua.

> CONCEITO-CHAVE A variação e a distribuição dos poligenes pode contribuir para a variação contínua em uma população.

3.5 Genes de organelas | Herança independente do núcleo

Até agora, consideramos como os genes nucleares se distribuem independentemente em virtude de seus *loci* em diferentes cromossomos. Entretanto, embora o núcleo contenha a maior parte dos genes de um organismo eucariótico, um subconjunto distinto e especializado do genoma é encontrado nas mitocôndrias e, em plantas, também nos cloroplastos. Esses subconjuntos são herdados independentemente do genoma nuclear e, assim, constituem um caso especial de herança independente, por vezes denominada herança extranuclear.

As mitocôndrias e os cloroplastos são organelas especializadas localizadas no citoplasma. Elas contêm pequenos cromossomos circulares que carreiam um subconjunto definido do genoma celular total. Os genes mitocondriais estão relacionados com a tarefa de produção de energia da mitocôndria, enquanto os genes dos cloroplastos são necessários para que o cloroplasto realize a sua função de fotossíntese. Entretanto, nenhuma organela é funcionalmente autônoma, tendo em vista que cada uma delas depende em grande parte dos genes nucleares para a sua função. O motivo pelo qual alguns dos genes necessários estão nas próprias organelas e outros estão no núcleo ainda é algo misterioso, que não será abordado aqui.

Outra peculiaridade dos genes das organelas é a grande quantidade de cópias presentes em uma célula. Cada organela tem muitas cópias por célula e, além disso, cada organela contém muitas cópias do seu cromossomo. Portanto, cada célula pode conter centenas ou milhares de cromossomos de organelas. Considere os cloroplastos, por exemplo, qualquer célula verde de uma planta apresenta muitos cloroplastos, e cada cloroplasto contém muitas moléculas idênticas de DNA circular, os assim denominados cromossomos de cloroplasto. Portanto, o

número de cromossomos de cloroplasto por célula pode ser da ordem de milhares e o número pode até mesmo variar um pouco de célula para célula. O DNA por vezes é observado acondicionado em estruturas suborganelares, denominadas *nucleoides*, que se tornam visíveis se usado um corante de ligação ao DNA (Figura 3.18). O DNA encontra-se dobrado dentro do nucleoide, mas não apresenta o tipo de helicoidização associada às histonas demonstrado pelos cromossomos nucleares. O mesmo arranjo é verdadeiro em relação ao DNA nas mitocôndrias. Por enquanto, presumiremos que todas as cópias de um cromossomo de organela dentro de uma célula são idênticas, mas posteriormente precisaremos relaxar essa presunção.

Célula demonstrando nucleoides dentro das mitocôndrias

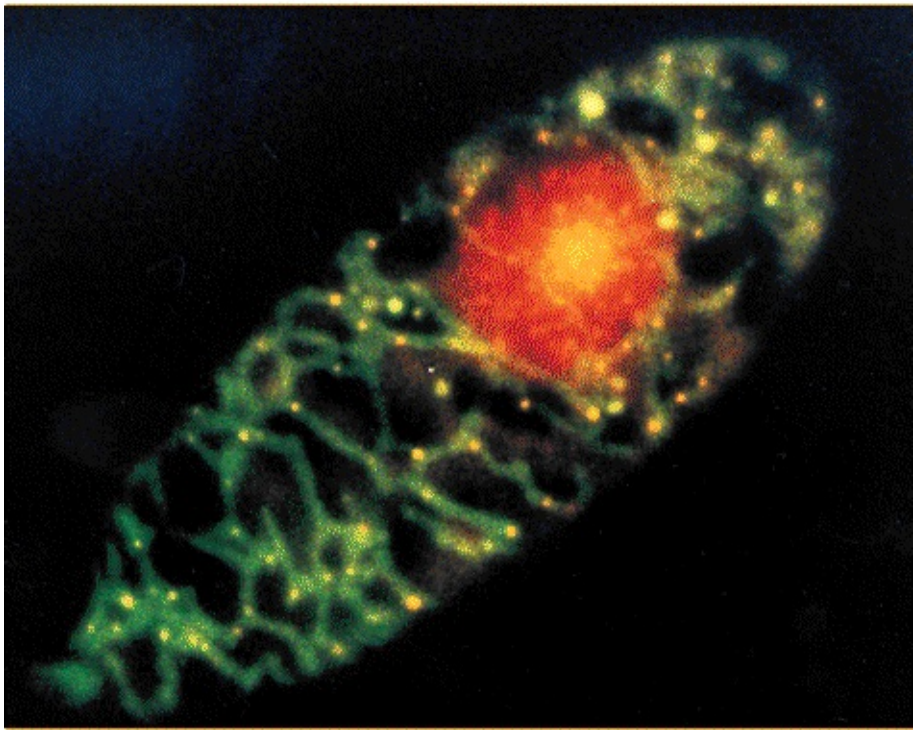


FIGURA 3.18 Coloração fluorescente de uma célula de *Euglena gracilis*. Com os corantes utilizados, o núcleo aparece em vermelho em virtude da fluorescência de grandes quantidades de DNA nuclear. As mitocôndrias fluorescem em verde e, dentro das mitocôndrias, as concentrações de DNA mitocondrial (nucleoides) fluorescem em amarelo. (De Y. Hayashi e K. Ueda, "The shape of mitochondria and the number of mitochondrial nucleoids during the cell cycle of *Euglena gracilis*", *J. Cell Sci.* 93, 1989, 565. Foto por The Company of Biologists Ltd.)

Muitos cromossomos de organelas atualmente foram sequenciados. Exemplos de tamanho e espaçamento relativo dos genes no **DNA mitocondrial (mtDNA)** e no **DNA de cloroplasto (cpDNA)** estão demonstrados na Figura 3.19. Os genes de organelas estão espaçados de modo muito próximo e, em alguns organismos, os genes de organelas podem conter segmentos não traduzidos denominados *íntrons*. Observe como a maior parte dos genes está relacionada com as reações químicas que ocorrem dentro da própria organela: a fotossíntese em cloroplastos e a fosforilação oxidativa em mitocôndrias.

Padrões de herança em organelas

Os genes de organelas demonstram seu próprio modo de herança especial, denominado **herança uniparental**: a progênie herda os genes de organelas exclusivamente de um dos genitores, mas não do outro. Na maior parte dos casos, o genitor é a mãe, um padrão denominado **herança materna**. Por que apenas a mãe? A resposta encontra-se no fato de que os cromossomos de organelas estão localizados no citoplasma e os gametas masculino e feminino não contribuem com o citoplasma igualmente para o zigoto. No que diz respeito aos genes nucleares, ambos os genitores contribuem igualmente para o zigoto. Entretanto, o ovócito contribui com a maior parte do citoplasma, enquanto o espermatozoide quase não contribui. Portanto, tendo em vista que as organelas estão localizadas no citoplasma, a genitora contribui com as organelas juntamente com o citoplasma, e essencialmente nenhum DNA de organela no zigoto é do genitor.

Algumas variantes fenotípicas são causadas por um alelo mutante de um gene de organela, e podemos utilizar esses mutantes para seguir os padrões da herança de organelas. Presumiremos temporariamente que o alelo mutante está presente em todas as cópias do cromossomo da organela, uma situação que, de fato, é observada com frequência. Em um cruzamento, o fenótipo variante será transmitido para a progênie se a variante utilizada for a genitora, mas não se for o genitor. Portanto, em geral, a herança citoplasmática demonstra o padrão a seguir:

Fêmea mutante × Macho do tipo selvagem → Toda a progênie mutante
Fêmea do tipo selvagem × Macho mutante → Toda a progênie do tipo

selvagem

De fato, esse padrão de herança é diagnóstico da herança de organelas nos casos em que a localização genômica de um alelo mutante não é conhecida.

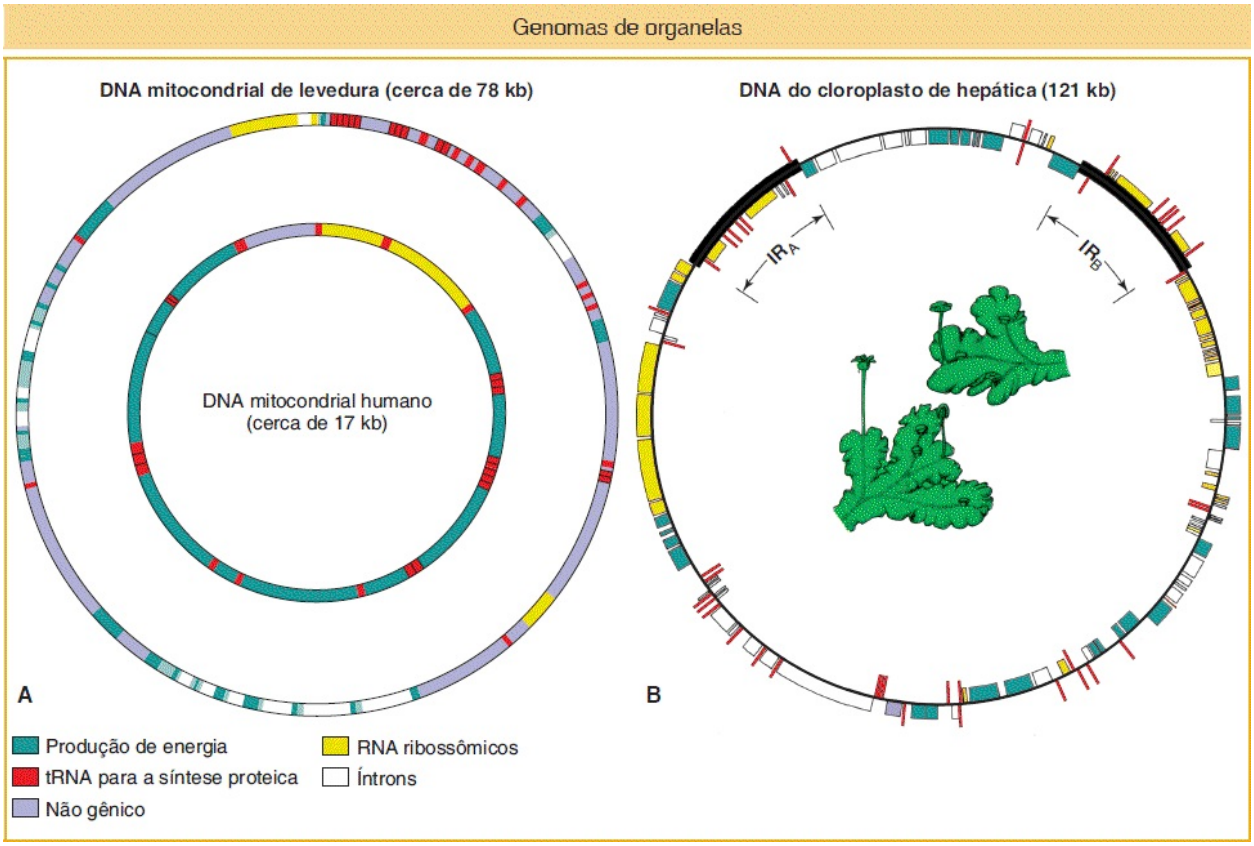


FIGURA 3.19 Mapas de DNA em relação a mitocôndrias e cloroplastos. Muitos dos genes de organelas codificam proteínas que realizam as funções de produção de energia destas organelas (*verdes*), enquanto outros (*vermelhos* e *laranja*) atuam na síntese proteica. A. Mapas de mtDNA de levedura e humano. (Observe que o mapa humano não está desenhado na mesma escala que o mapa de levedura.) B. O genoma de cloroplasto de 121 kb da planta *Marchantia polymorpha*. Os genes demonstrados no mapa estão transcritos em sentido horário e os do lado de fora estão transcritos em sentido anti-horário. IR_A e IR_B indicam repetições invertidas. O desenho superior no centro do mapa ilustra uma planta *Marchantia* do sexo masculino; o desenho inferior ilustra uma planta do sexo feminino. (Dados de K. Umesono e H. Ozeki, *Trends Genet.* 3, 1987.)

A herança materna pode ser claramente demonstrada em determinados mutantes de fungos. Por exemplo, no fungo *Neurospora*, um mutante denominado *poky* apresenta um fenótipo de crescimento lento. Pode-se cruzar *Neurospora* de tal

modo que um genitor atue como a genitora materna, contribuindo com o citoplasma (ver Figura 3.9). Os resultados dos cruzamentos recíprocos a seguir sugerem que o gene mutante está localizado nas mitocôndrias (os fungos não apresentam cloroplastos):

Fêmea *poky* × Macho do tipo selvagem →
Toda a progênie *poky*
Fêmea do tipo selvagem × Macho *poky* →
Toda a progênie do tipo selvagem

O sequenciamento demonstrou que o fenótipo *poky* é causado por uma mutação de um gene de RNA ribossômico no mtDNA. A sua herança está demonstrada diagramaticamente na Figura 3.20. O cruzamento inclui uma diferença alélica (*ad* e ad^{+}) em um gene nuclear adicional ao *poky*; observe como a herança mendeliana do gene nuclear é independente da herança materna do fenótipo *poky*.

> CONCEITO-CHAVE Os fenótipos variantes causados por mutações no DNA de organelas citoplasmáticas em geral são herdados de modo materno e independente dos padrões mendelianos demonstrados pelos genes nucleares.

Segregação citoplasmática

Em alguns casos, as células contêm misturas de organelas mutantes e normais. Essas células são denominadas *cytohets* ou *heteroplasmons*. Nessas misturas, pode ser detectado um tipo de **segregação citoplasmática**, no qual os próprios dois tipos se distribuem em diferentes células-filhas. O processo muito provavelmente tem origem na divisão aleatória das organelas múltiplas durante a divisão celular. As plantas fornecem um bom exemplo. Muitos casos de folhas brancas são causados por mutações nos genes dos cloroplastos que controlam a produção e a deposição do pigmento verde clorofila. Tendo em vista que a clorofila é necessária para que uma planta viva, esse tipo de mutação é letal, e as plantas com folhas brancas não podem ser obtidas para cruzamentos

experimentais. Entretanto, algumas plantas são variegadas, contendo manchas verdes e brancas, e essas plantas são viáveis. Portanto, as plantas variegadas proporcionam um modo para demonstrar a segregação citoplasmática.

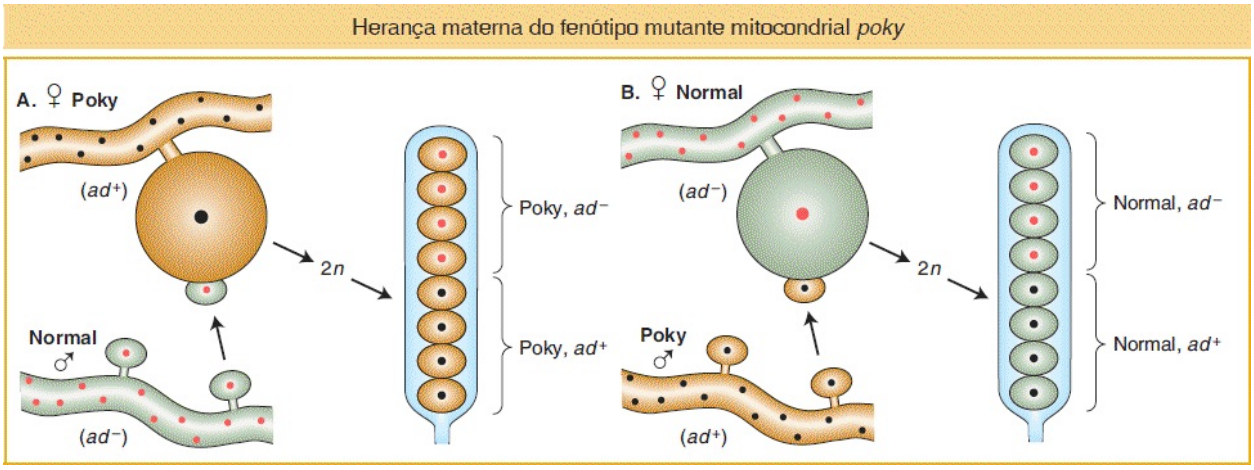


FIGURA 3.20 Cruzamentos recíprocos de *Neurospora poky* e do tipo selvagem produzem resultados diferentes, tendo em vista que um genitor diferente contribui com o citoplasma. A genitora contribui com a maior parte do citoplasma das células da progênie. O *sombreamento em marrom* representa o citoplasma com mitocôndrias que contêm a mutação *poky* e o *sombreamento em verde* representa o citoplasma com mitocôndrias do tipo selvagem. Observe que toda a progênie na parte **A** é *poky*, enquanto toda a progênie na parte **B** é normal. Portanto, ambos os cruzamentos demonstram herança materna. O gene nuclear com os alelos ad^+ (*preto*) e ad^- (*vermelho*) é utilizado para ilustrar a segregação dos genes nucleares na proporção mendeliana de 1:1 esperada para este organismo haploide.

A planta maravilha na Figura 3.21 exibe um fenótipo comumente observado de folhas e ramos variegados, que demonstra a herança de um alelo mutante de um gene de cloroplasto. O alelo mutante faz com que os cloroplastos sejam brancos; por sua vez, a cor dos cloroplastos determina a cor das células e, portanto, a cor dos ramos compostos por aquelas células. Os ramos variegados são mosaicos de células todas verdes e todas brancas. As flores podem se desenvolver em ramos verdes, brancos ou variegados, e os genes de cloroplasto nas células de uma flor são aqueles do ramo no qual ela cresce. Portanto, em um cruzamento (Figura 3.22), o gameta materno dentro da flor (o ovócito) determina a cor das folhas e dos ramos das plantas da progênie. Por exemplo, se um ovócito for de uma flor em um ramo verde, toda a progênie será verde, independentemente da origem do pólen. Um ramo branco apresentará cloroplastos brancos, e as plantas da

progênie resultantes serão brancas. (Em virtude da letalidade, os descendentes brancos não viveriam além do estágio de muda.)

Os zigotos variegados (parte inferior da Figura 3.22) demonstram segregação citoplasmática. Essa progênie variegada tem origem em ovócitos que são *cytohets*. Curiosamente, quando um referido zigoto se divide, os cloroplastos brancos e verdes com frequência segregam; ou seja, eles próprios se dividem em células separadas, produzindo os setores verdes e brancos distintos que causam a variegação nos ramos. Portanto, aqui está uma demonstração direta da segregação citoplasmática.

Tendo em vista que uma célula é uma população de moléculas de organelas, como é possível a obtenção de uma célula mutante "pura", que contém apenas cromossomos mutantes? Muito provavelmente, mutantes puros são criados em células assexuadas, como segue. As variantes surgem por mutação de um único gene em um único cromossomo. Em seguida, em alguns casos, o cromossomo que contém a mutação apresenta aumento aleatório da frequência na população dentro da célula. Esse processo é denominado *deriva genética aleatória*. Uma célula que é um *cytohet* pode apresentar, digamos, 60% de cromossomos *A* e 40% de cromossomos *a*. Quando essa célula se divide, por vezes todos os cromossomos *A* dirigem-se para uma célula-filha, e todos os cromossomos *a* para a outra (novamente, ao acaso). Com mais frequência, essa separação requer diversas gerações subsequentes de divisão celular para ser completada (Figura 3.23). Portanto, como resultado desses eventos ao acaso, ambos os alelos são expressos em diferentes células-filhas, e essa separação continuará nos descendentes dessas células. Observe que a segregação citoplasmática não é um processo mitótico; ela ocorre na divisão das células assexuadas, mas não está relacionada com a mitose. Em cloroplastos, a segregação citoplasmática é um mecanismo comum para a produção de plantas variegadas (verdes e brancas), conforme já mencionado. Em mutantes de fungos, tal como o mutante *poky* de *Neurospora*, a mutação original em uma molécula de mtDNA necessariamente acumulou e sofreu segregação citoplasmática para produzir a linhagem que expressa os sintomas *poky*.

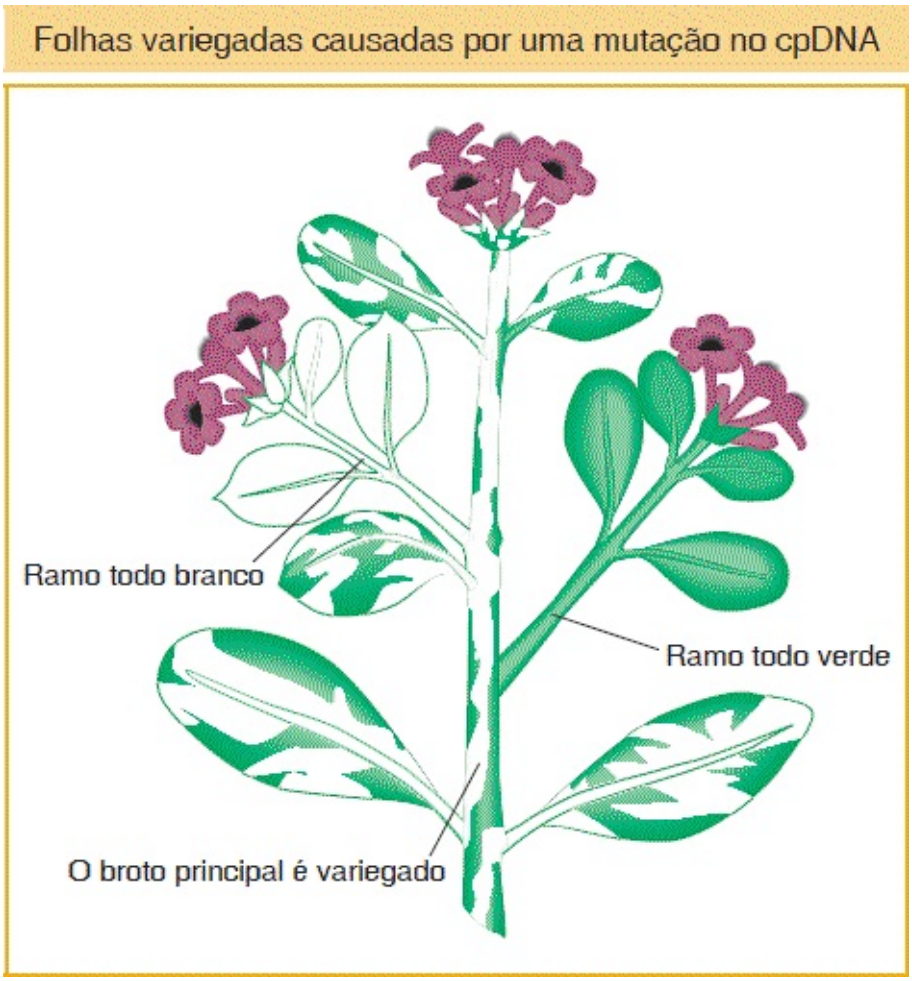


FIGURA 3.21 Variegação das folhas na *Mirabilis jalapa*, a planta maravilha. As flores podem se formar em qualquer ramo (*variegado*, *verde* ou *branco*) e ser utilizadas em cruzamentos.

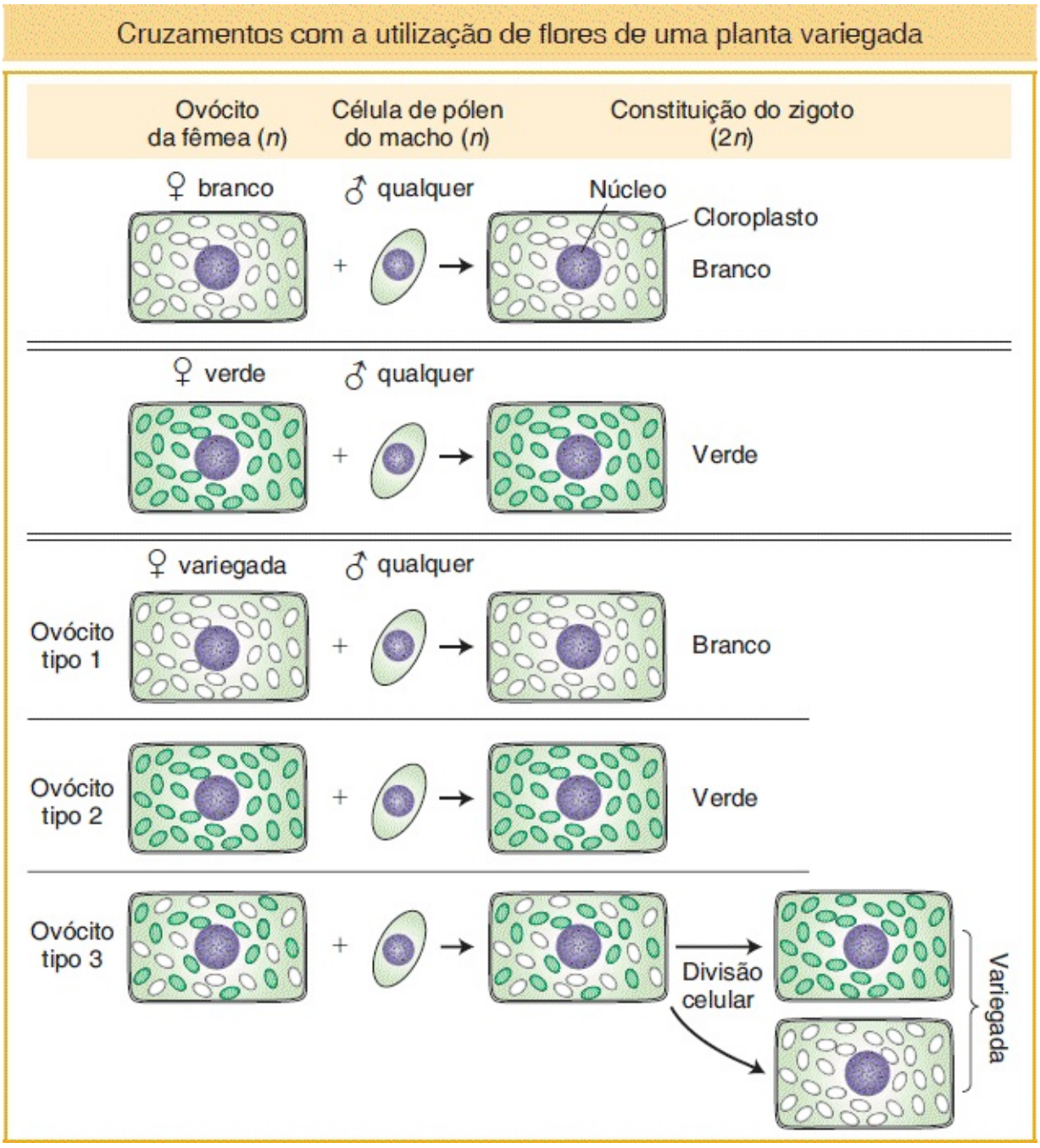


FIGURA 3.22 Os resultados dos cruzamentos da *Mirabilis jalapa* podem ser explicados pela herança autônoma dos cloroplastos. As esferas grandes e escuras representam os núcleos. Os corpos menores representam os cloroplastos, sejam verdes ou brancos. Presume-se que cada ovócito contenha muitos cloroplastos, e presume-se que cada célula de pólen não contenha cloroplastos. Os primeiros dois cruzamentos exibem herança materna estrita. Entretanto, se o ramo materno for variegado, podem resultar três tipos de zigotos, dependendo de o ovócito conter apenas cloroplastos brancos, apenas verdes, ou ambos, verdes e brancos. No último caso, o zigoto resultante pode produzir ambos os tecidos verde e branco e, assim, resulta uma planta variegada.

CONCEITO-CHAVE As populações de organelas que contêm misturas de dois cromossomos geneticamente distintos com frequência demonstram segregação dos dois tipos nas células-filhas na divisão celular. Esse processo é denominado segregação citoplasmática.

Em determinados sistemas especiais, tal como em fungos e em algas, foram obtidos *cytohets* que são "di-híbridos" (digamos, *AB* em um cromossomo de organela e *ab* em outro). Nos referidos casos, podem ocorrer processos raros semelhantes ao *crossing*, mas tal ocorrência tem de ser considerada um fenômeno genético menor.

CONCEITO-CHAVE Alelos em cromossomos de organelas:

1. Em cruzamentos sexuados, são herdados de apenas um genitor (em geral do sexo feminino) e, portanto, não demonstram proporções de segregação do tipo nuclear que os genes nucleares demonstram.
2. Em células assexuadas, podem demonstrar segregação citoplasmática.
3. Em células assexuadas, ocasionalmente há processos análogos ao *crossing over*.

Mutações citoplasmáticas em humanos

Existem mutações citoplasmáticas em humanos? Alguns heredogramas humanos demonstram a transmissão de distúrbios raros apenas por meio das mulheres e nunca por meio dos homens. Esse padrão sugere fortemente a herança citoplasmática e aponta para uma mutação no mtDNA como o motivo para o fenótipo. A doença MERFF (epilepsia mioclônica com fibras vermelhas rotas) é um referido fenótipo, que resulta de uma alteração de uma base no mtDNA. É uma doença que afeta os músculos, mas os sintomas também incluem distúrbios oculares e auditivos. Outro exemplo é a síndrome de Kearns-Sayre, uma constelação de sintomas que afeta olhos, coração, músculos e cérebro, que é causada pela perda de parte do mtDNA. Em alguns desses casos, as células de uma pessoa afetada contêm misturas de cromossomos normais e mutantes, e as

proporções de cada um que são transmitidas para a progênie podem variar conforme resultado da segregação citoplasmática. As proporções em uma pessoa também podem variar em diferentes tecidos ou ao longo do tempo. O acúmulo de determinados tipos de mutações mitocondriais ao longo do tempo tem sido proposto como uma possível causa de envelhecimento.

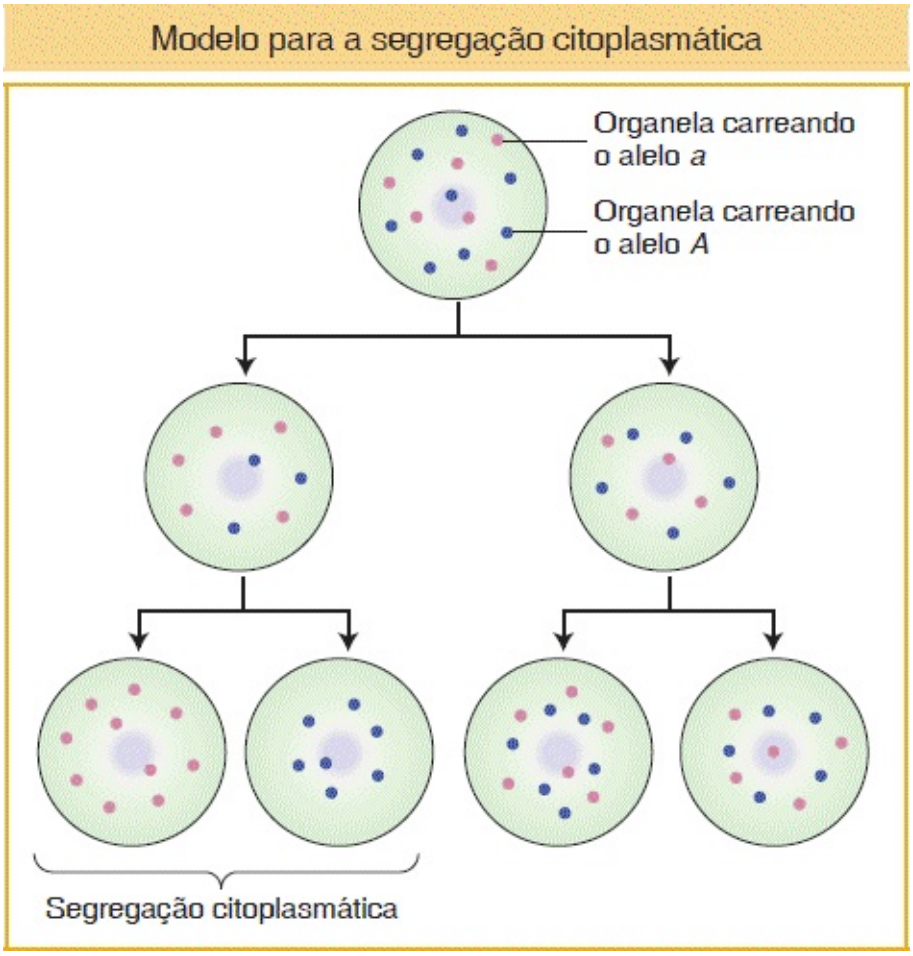


FIGURA 3.23 Ao acaso, organelas geneticamente distintas podem segregar para células separadas em diversas divisões celulares sucessivas. Os *pontos vermelhos* e *azuis* representam organelas geneticamente distinguíveis, tais como mitocôndrias com e sem uma mutação.

A Figura 3.24 demonstra algumas das mutações em genes mitocondriais humanos que podem levar à doença quando, por meio de deriva aleatória e segregação citoplasmática, aumentam em frequência até uma tal medida que a função celular é comprometida. A herança de uma doença mitocondrial humana

está demonstrada na Figura 3.25. Observe que a condição sempre é transmitida para a descendência pelas mães, e nunca pelos pais. Ocasionalmente, a mãe terá um filho não afetado (não demonstrado), provavelmente em virtude da segregação citoplasmática no tecido que forma o gameta.

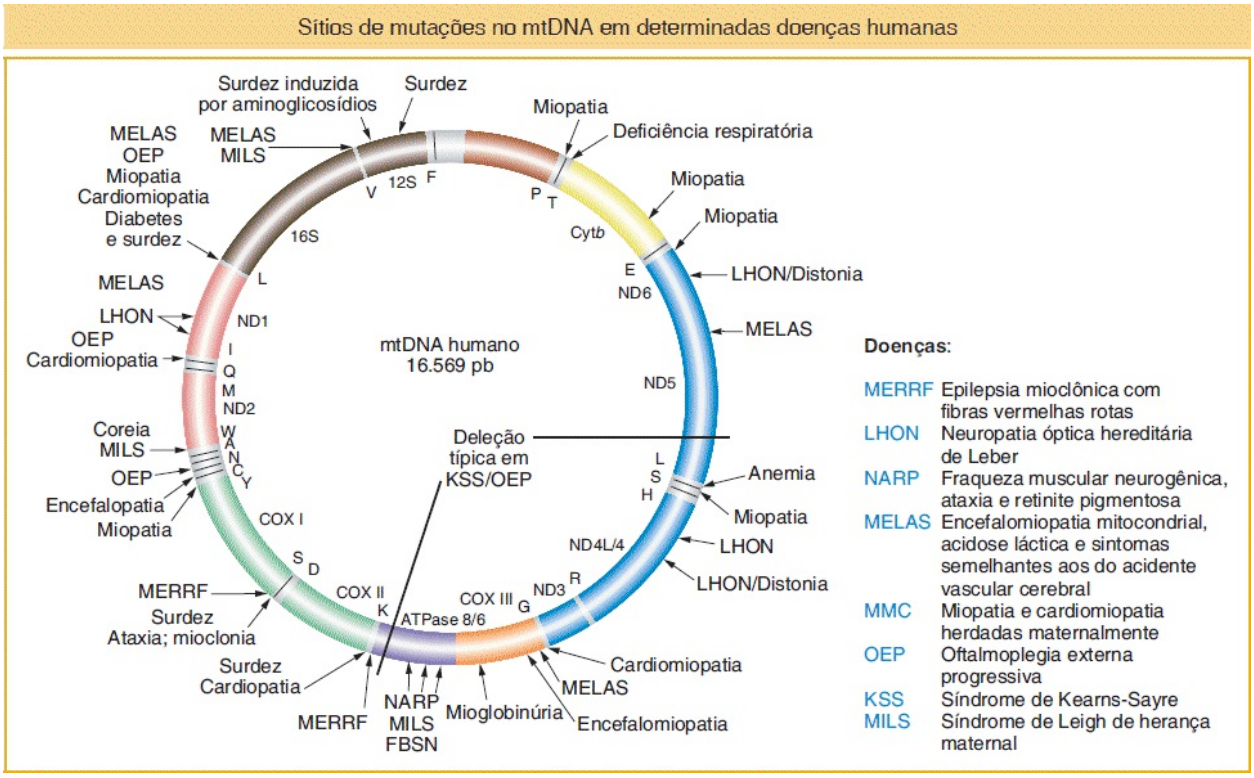


FIGURA 3.24 Este mapa do mtDNA humano demonstra *loci* de mutações que levam a citopatias. Os genes de RNA transportador estão representados por abreviações de aminoácidos de letra única: ND = NADH desidrogenase; COX = Citocromo C oxidase; e 12S e 16S fazem referência aos RNA ribossômicos. (Dados de S. DiMauro *et al.*, "Mitochondria in Neuromuscular Disorders", *Biochim. Biophys. Acta* 1366, 1998, 199-210.)

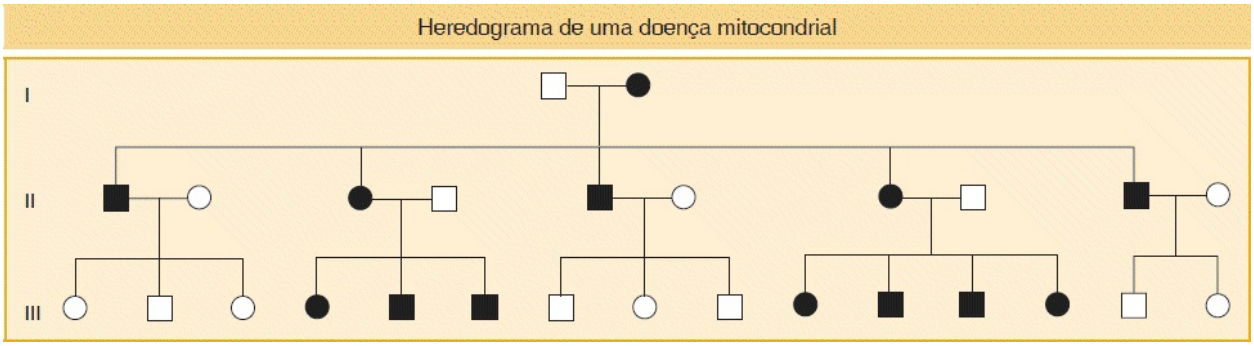


FIGURA 3.25 Este heredograma demonstra que uma doença mitocondrial humana é herdada apenas da mãe.

mtDNA em estudos evolutivos

As diferenças e as semelhanças das sequências de mtDNA homólogas entre as espécies foram utilizadas extensivamente para construir árvores evolutivas. Além disso, foi possível introduzir alguns organismos extintos em árvores evolutivas com a utilização de sequências de mtDNA obtidas dos remanescentes de organismos extintos, tais como peles e ossos em museus. O mtDNA evolui relativamente rápido e assim essa abordagem tem sido mais útil no traçado da evolução recente, tal como a evolução de humanos e outros primatas. Um achado-chave é que a "raiz" da árvore do mtDNA humano está na África, sugerindo que o *Homo sapiens* teve origem na África e a partir de lá se dispersou por todo o mundo (ver Capítulo 18).

RESUMO

A pesquisa genética e o cultivo de plantas e a criação de animais com frequência necessitam da síntese de genótipos que são combinações complexas de alelos de diferentes genes. Os referidos genes podem estar no mesmo cromossomo ou em cromossomos diferentes; o último é o assunto principal deste capítulo.

No caso mais simples — um di-híbrido em relação ao qual os dois pares de genes estão em pares de cromossomos diferentes — cada par de genes demonstra segregação igual na meiose, conforme previsto pela primeira lei de Mendel. Tendo em vista que no núcleo as fibras do fuso se ligam aleatoriamente aos centrômeros na meiose, os dois pares de genes são separados independentemente dentro dos produtos meióticos. Esse princípio de segregação independente é denominado segunda lei de Mendel, tendo em vista que Mendel foi o primeiro a observá-lo. A partir de um di-híbrido *A*/*a* ; *B*/*b*, são produzidos quatro genótipos de produtos meióticos, *A*; *B*, *A*; *b*, *a*; *B* e *a*; *b*, todos a uma frequência igual de 25% cada. Portanto, em um cruzamento-teste de um di-híbrido com um recessivo duplo, as proporções fenotípicas da progênie também são de 25% (uma proporção de 1:1:1:1). Se um referido di-híbrido é autofecundado, as classes

fenotípicas na progênie são *A*/—; *B*/—, *A*/—; *b*/*b*, *a*/*a*; *B*/— e *a*/*a*; *b*/*b*. As proporções de 1:1:1:1 e 9:3:3:1 são ambas diagnósticas da segregação independente.

Genótipos mais complexos, compostos por genes de distribuição independente, podem ser tratados como extensões do caso de segregação monogênica. As proporções genotípicas, fenotípicas ou gaméticas em geral são calculadas por meio da aplicação da regra do produto — ou seja, por meio da multiplicação das proporções relevantes de genes individuais. A probabilidade de ocorrência de qualquer uma das diversas categorias da progênie é calculada por meio da aplicação da regra da soma — ou seja, por meio da adição de suas probabilidades individuais. No tipo mnemônico, a regra do produto lida com "A E B", enquanto a regra da soma lida com "A′ OU A″". O teste do χ^2 pode ser utilizado para testar se as proporções observadas das classes na análise genética estão conforme as expectativas de uma hipótese genética, tal como uma hipótese de herança monogênica ou digênica. Se um valor de probabilidade inferior a 5% for calculado, a hipótese tem de ser rejeitada.

Gerações sequenciais de autocruzamento aumentam as proporções de homozigotos, de acordo com os princípios da segregação igual e da segregação independente (se os genes estiverem em cromossomos diferentes). Portanto, o autocruzamento é utilizado para criar linhagens puras complexas com combinações de mutações desejáveis.

A distribuição independente dos cromossomos na meiose pode ser observada citologicamente por meio da utilização de pares de cromossomos heteromórficos (aqueles que demonstram uma diferença estrutural). Os cromossomos X e Y são um referido caso, mas outros casos mais raros podem ser observados e utilizados para essa demonstração. A distribuição independente dos genes no nível de meiócitos únicos pode ser observada nos fungos ascomicetos, tendo em vista que os ascos demonstram os dois tipos alternativos de segregações em frequências iguais.

Uma das principais funções da meiose é produzir recombinantes, novas combinações de alelos dos genótipos haploides que se uniram para formar o meiócito. A distribuição independente é a principal fonte de recombinantes. Em

um cruzamento-teste de di-híbridos que demonstra distribuição independente, a frequência de recombinantes será de 50%.

Características métricas, tais como a intensidade da cor, demonstram uma distribuição contínua em uma população. Distribuições contínuas podem ter por base uma variação ambiental, alelos variantes de genes múltiplos ou uma combinação de ambos. Um modelo genético simples propõe que alelos ativos de diversos genes (denominados poligenes) contribuem de modo mais ou menos aditivo para o caractere métrico. Em uma análise da progênie a partir do autocruzamento de um indivíduo multiplamente heterozigoto, o histograma que demonstra a proporção de cada fenótipo se aproxima de uma curva sinusoidal típica da variação contínua.

Os pequenos subconjuntos do genoma observados em mitocôndrias e cloroplastos são herdados independentemente do genoma nuclear. Mutantes nesses genes de organelas com frequência demonstram herança materna, juntamente com o citoplasma, que é a localização dessas organelas. Em citoplasmas geneticamente mistos (*cytohets*), os dois genótipos (digamos, do tipo selvagem e mutante) com frequência se separam em diferentes células-filhas por meio de um processo pouco compreendido denominado segregação citoplasmática. A mutação mitocondrial em seres humanos resulta em doenças que demonstram segregação citoplasmática nos tecidos corporais e herança materna em um cruzamento.

TERMOS-CHAVE

cruzamento di-híbrido
di-híbrido
DNA de cloroplasto (cpDNA)
DNA mitocondrial (mtDNA)
herança materna
herança uniparental
lei da segregação independente (segunda lei de Mendel)
locus de traço quantitativo (QTL)

poligene (*locus* de traço quantitativo)
recombinação
recombinação meiótica
recombinante
regra da soma
regra do produto
segregação citoplasmática
segregação (distribuição) independente
teste do qui-quadrado
vigor híbrido

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Duas moscas *Drosophila* que apresentavam asas normais (transparentes e longas) foram cruzadas. Na progênie, apareceram dois fenótipos novos: asas pardas (que apresentam um aspecto semiopaco) e asas aparadas (com extremidades quadradas). A progênie foi como segue:

Fêmeas	Machos
179 transparentes e longas	92 transparentes e longas
58 transparentes e aparadas	89 pardas e longas
	28 transparentes e aparadas
	31 pardas e aparadas

a. Forneça uma explicação cromossômica para esses resultados, demonstrando os genótipos cromossômicos dos genitores e de todas as classes de progênie sob o seu modelo.
b. Desenhe um teste para o seu modelo.

Solução

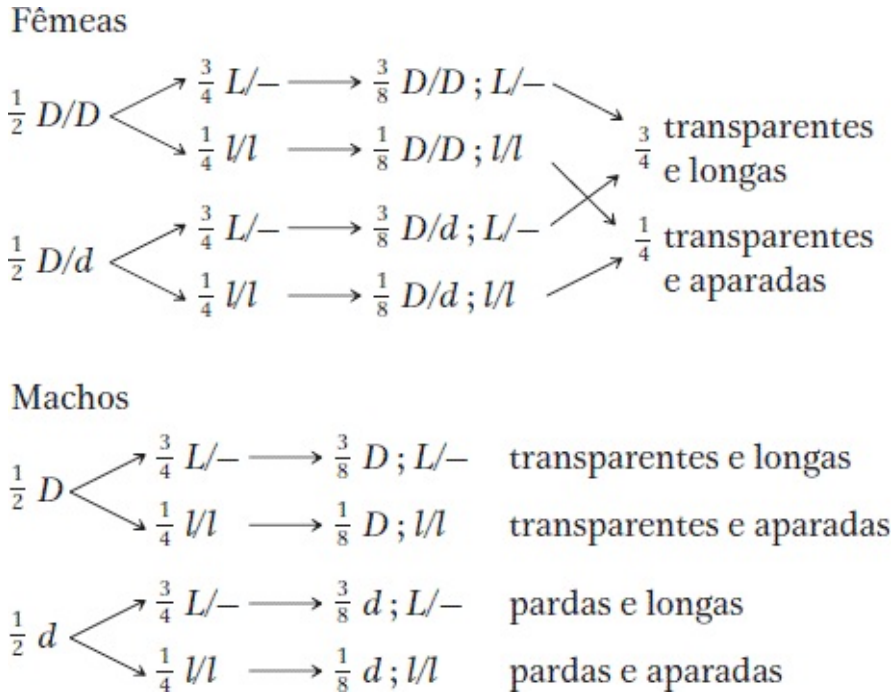
a. O primeiro passo é declarar quaisquer características interessantes dos dados. A primeira característica surpreendente é o aparecimento de dois fenótipos novos. Encontramos o fenômeno no Capítulo 2, onde ele foi explicado como alelos recessivos mascarados por seus correspondentes dominantes. Assim, primeiramente poderíamos supor que uma ou ambas as moscas genitoras apresentam alelos recessivos de dois genes diferentes. Essa inferência é fortalecida pela observação de que parte da progênie expressa apenas um dos fenótipos novos. Se os novos fenótipos sempre apareceram em conjunto, podemos supor que o mesmo alelo recessivo determina ambos.

Entretanto, a outra característica surpreendente dos dados, que não conseguimos explicar por meio da utilização dos princípios mendelianos do Capítulo 2, é a diferença óbvia entre os sexos: embora existam quantidades aproximadamente iguais de machos e fêmeas, os machos se situam dentro de quatro classes fenotípicas, mas as fêmeas constituem apenas duas. Esse fato deve sugerir imediatamente algum tipo de herança ligada ao sexo. Quando estudamos os dados, observamos que os fenótipos longos e aparados estão se segregando em ambos machos e fêmeas, mas apenas os machos apresentam o fenótipo pardo. Essa observação sugere que a herança da transparência das asas difere da herança do formato das asas. Primeiramente, longas e aparadas são observadas a uma proporção de 3:1 em ambos machos e fêmeas. Essa proporção pode ser explicada se ambos os genitores forem heterozigotos em relação a um gene autossômico; podemos representá-los como *L*/*l*, em que *L* faz referência a longas e *l* faz referência a aparadas.

Tendo realizado essa análise parcial, observamos que apenas a herança da transparência das asas está associada ao sexo. A possibilidade mais óbvia é de que os alelos em relação a transparente (*D*) e parda (*d*) estejam no cromossomo X, pois vimos no Capítulo 2 que a localização dos genes nesse cromossomo proporciona padrões de herança correlacionados com o sexo. Se essa sugestão for verdadeira, a fêmea genitora tem de ser aquela que porta o alelo *d*, tendo em vista que, se o macho apresentasse o *d*, ele teria sido pardo, sendo que nos é informado que ele apresentava asas transparentes. Portanto, a genitora do sexo feminino seria *D*/*d* e o macho, *D*. Vejamos se essa sugestão funciona: se ela for

verdadeira, toda a progênie do sexo feminino herdaria o alelo *D* de seu pai e, assim, todas seriam aladas e transparentes, conforme foi observado. Metade dos filhos seria *D* (transparentes) e metade, *d* (pardas), o que também foi observado.

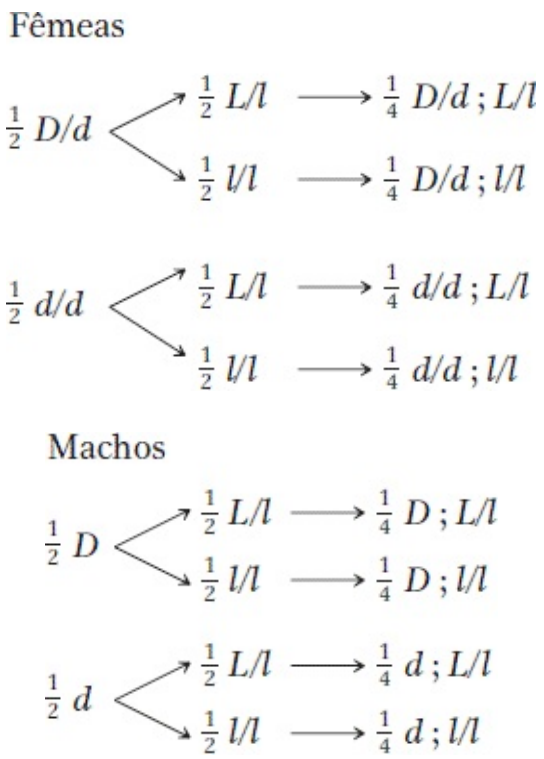
Assim, em geral, podemos representar a genitora como *D*/*d*; *L*/*l* e o genitor como *D*; *L*/*l*. Então, a progênie seria:



b. Em geral, um bom modo de testar um referido modelo é realizar um cruzamento e prever o desfecho. Mas qual cruzamento? Devemos prever algum tipo de proporção na progênie e, assim, é importante realizar um cruzamento a partir do qual uma proporção fenotípica única possa ser esperada. Observe que a utilização de uma descendente como genitora não atenderia as nossas necessidades: não podemos dizer, a partir da observação do fenótipo de quaisquer dessas fêmeas, qual é o seu genótipo. Uma fêmea com asas transparentes pode ser *D*/*D* ou *D*/*d*, e uma com asas longas pode ser *L*/*L* ou *L*/*l*. Seria bom cruzar a fêmea parental do cruzamento original com um filho pardo e aparado, tendo em vista que os genótipos totais de ambos estão especificados sob o modelo que criamos. De acordo com o nosso modelo, esse cruzamento é:

D/*d*; *L*/*l* × *d*; *l*/*l*

A partir desse cruzamento, prevemos:



Problema resolvido 2. Considere três ervilhas amarelas e lisas, rotuladas A, B e C. Cada uma foi cultivada até uma planta e cruzada com uma planta cultivada a partir de uma ervilha verde rugosa. Exatamente 100 ervilhas originárias de cada cruzamento foram separadas em classes fenotípicas, como segue:

A:	51 amarelas e lisas
	49 verdes e lisas
B:	100 amarelas e lisas
C:	24 amarelas e lisas
	26 amarelas e rugosas
	25 verdes e lisas

25 verdes e rugosas

Quais eram os genótipos de A, B e C? (Utilize símbolos gênicos de sua própria escolha; assegure-se de definir cada um deles.)

Solução

Observe que cada um dos cruzamentos é:

$$\text{Amarela e lisa} \times \text{Verde e rugosa} \rightarrow \text{Progênie}$$

Tendo em vista que A, B e C foram todas cruzadas com a mesma planta, todas as diferenças entre as três populações de progênie têm de ser atribuíveis às diferenças nos genótipos subjacentes a A, B e C.

Você pode relembrar muito a respeito dessas análises a partir do capítulo, o que é bom, mas vejamos quanto podemos deduzir a partir dos dados. O que dizer a respeito da dominância? O cruzamento-chave para a dedução da dominância é B. Aqui, o padrão de herança é:

$$\text{Amarela e lisa} \times \text{Verde e rugosa} \rightarrow \text{Todas amarelas e lisas}$$

Assim, amarela e lisa têm de ser fenótipos dominantes, tendo em vista que a dominância é literalmente definida pelo fenótipo de um híbrido. Agora sabemos que o genitor verde e rugoso utilizado em cada cruzamento tem de ser totalmente recessivo; temos uma situação muito conveniente, tendo em vista que ela significa que cada cruzamento é um cruzamento-teste, que em geral é o tipo de cruzamento mais informativo.

Voltando para a progênie de A, observamos uma proporção de 1:1 em relação a amarelas e verdes. Essa proporção é uma demonstração da primeira lei de Mendel (segregação igual) e demonstra que, em relação à característica da cor, o cruzamento foi obrigatoriamente heterozigoto × homozigoto recessivo. Deixando *Y* representar amarela e *y* representar verde, temos:

$$Y/y \times y/y \rightarrow Y/y \text{ (amarelas)} \rightarrow y/y \text{ (verdes)}$$

Em relação ao formato, tendo em vista que toda a progênie é lisa, o cruzamento foi necessariamente homozigoto dominante × homozigoto recessivo. Deixando *R* representar lisas e *r* representar rugosas, temos:

$$R/R \times r/r \rightarrow R/r \text{ (lisas)}$$

Combinando as duas características, temos:

$$Y/y;\ R/R \times y/y;\ r/r \rightarrow Y/y;\ R/r,\ y/y;\ R/r$$

Agora, o cruzamento B está claro e tem de ser:

$$Y/Y;\ R/R \times y/y;\ r/r \rightarrow Y/y;\ r/r$$

tendo em vista que qualquer heterozigosidade na ervilha B teria dado origem a diversos fenótipos na progênie, não apenas um.

O que dizer sobre C? Aqui, observamos uma proporção de 50 amarelas:50 verdes (1:1) e uma proporção de 49 lisas:51 rugosas (também 1:1). Assim, ambos os genes na ervilha C são obrigatoriamente heterozigotos, e o cruzamento C foi:

$$Y/y;\ R/r \times y/y;\ r/r$$

que é uma boa demonstração da segunda lei de Mendel (distribuição independente de genes diferentes).

Como um geneticista teria analisado esses cruzamentos? Basicamente, do mesmo modo que acabamos de fazer, porém com menos etapas de permeio. Possivelmente, algo como isto: “amarela e lisa dominante; segregação monogênica em A; B homozigoto dominante; segregação independente de dois genes em C.”

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Com a utilização da Tabela 3.1, responda as questões a seguir a respeito

dos valores de probabilidade (anteriormente):

a. Se χ^2 for calculado como sendo 17 com 9 gl, qual é o valor aproximado da probabilidade?

b. Se χ^2 for 17 com 6 gl, qual é o valor da probabilidade?

c. Qual tendência (“regra”) você observa nos dois cálculos anteriores?

2. Inspecione a Figura 3.8: qual estágio meiótico é responsável por gerar a segunda lei de Mendel?

3. Na Figura 3.9:

a. identifique os núcleos diploides.

b. identifique qual parte da figura ilustra a primeira lei de Mendel.

4. Inspecione a Figura 3.10: qual seria o desfecho na óctade se em raras ocasiões um núcleo da divisão mitótica pós-meiótica do núcleo 2 passasse para um núcleo da divisão mitótica pós-meiótica do núcleo 3? Como você poderia medir a frequência de um referido evento raro?

5. Na Figura 3.11, se os genótipos de entrada fossem a · B e A · b, quais seriam os genótipos de cor azul?

6. Na progênie observada na Figura 3.13, quais são as origens dos cromossomos de cor azul-escura, azul-clara e azul muito clara?

7. Na Figura 3.17, em qual barra do histograma o genótipo $R_1/r_1 \cdot R_2/R_2 \cdot r_3/r_3$ seria encontrado?

8. Ao examinar a Figura 3.19, qual você acredita que seja o principal motivo para a diferença no tamanho do mtDNA de uma levedura e humano?

9. Na Figura 3.20, qual cor é utilizada para indicar o citoplasma que contém mitocôndrias do tipo selvagem?

10. Na Figura 3.21, quais seriam os tipos de folhas da progênie da flor apical (parte superior)?

11. A partir do heredograma na Figura 3.25, qual princípio você pode deduzir a respeito da herança da doença mitocondrial a partir dos pais afetados?

PROBLEMAS BÁSICOS

12. Presuma que ocorre segregação independente e inicie com uma planta que é di-híbrida *A*/*a*; *B*/*b*:

a. Que proporção fenotípica é produzida a partir do autocruzamento?

b. Que proporção genotípica é produzida a partir do autocruzamento?

c. Que proporção fenotípica é produzida a partir do seu cruzamento-teste?

d. Que proporção genotípica é produzida a partir do seu cruzamento-teste?

13. A mitose normal ocorre em uma célula diploide de genótipo *A*/*A*; *B*/*b*. Quais dos genótipos a seguir poderiam representar as possíveis células-filhas?

a. *A*; *B*

b. *a*; *b*

c. *A*; *b*

d. *a*; *B*

e. *A*/*A*; *B*/*B*

f. *A*/*a*; *B*/*b*

g. *a*/*a*; *b*/*b*

14. Em um organismo diploide de $2n = 10$, presuma que você consiga rotular todos os centrômeros derivados da genitora e todos os centrômeros derivados do genitor. Quando esse organismo produz gametas, quantas combinações de centrômeros rotulados dos sexos masculino e feminino são possíveis nos gametas?

15. Demonstrou-se que quando um fino feixe de luz é direcionado para um núcleo, a quantidade de luz absorvida é proporcional ao conteúdo de DNA da célula. Com a utilização desse método, o DNA nos núcleos de diversos tipos de células diferentes em uma planta do milho foi comparado. Os números a seguir representam as quantidades relativas de DNA nesses diferentes tipos de células:

0,7, 1,4, 2,1, 2,8 e 4,2

Que células poderiam ter sido utilizadas para essas medições? (Nota: em plantas, a parte do endosperma da semente com frequência é triploide, 3*n*.)

16. Desenhe uma mitose haploide do genótipo a^+; *b*.

17. No musgo, os genes *A* e *B* são expressos apenas no gametófito. Possibilita-se que um esporófito de genótipo *A*/*a*; *B*/*b* produza gametófitos.
a. Qual proporção dos gametófitos será *A*; *B*?
b. Se a fertilização for aleatória, qual proporção de esporófitos na próxima geração será *A*/*a*; *B*/*b*?

18. Quando uma célula de genótipo *A*/*a*; *B*/*b*; *C*/*c* que apresenta todos os genes em pares de cromossomos separados se divide mitoticamente, quais são os genótipos das células-filhas?

19. Na levedura haploide *Saccharomyces cerevisiae*, os dois tipos reprodutivos são conhecidos como MATa e MATα. Você cruza uma linhagem roxa (ad^{-}) do tipo de cruzamento *a* e uma linhagem branca (ad^+) do tipo de cruzamento α. Se ad^- e ad^+ são alelos de um gene, e *a* e α são alelos de um gene herdado independentemente em um par de cromossomos separados, qual progênie você espera obter? Em quais proporções?

20. Em camundongos, o nanismo é causado por um alelo recessivo ligado ao X, e a pelagem rosa é causada por um alelo dominante autossômico (a pelagem normalmente é amarronzada). Se uma fêmea anã de uma linhagem pura for cruzada com um macho rosa de uma linhagem pura, quais serão as proporções fenotípicas na F_1 e na F_2 em cada sexo? (Invente e defina os seus próprios símbolos gênicos.)

21. Suponha que você descobre duas interessantes anormalidades citológicas *raras* no cariótipo de um homem. (Um cariótipo é o complemento de cromossomo visível total.) Existe uma parte extra (satélite) em *um* dos cromossomos do par 4 e existe um padrão anormal de coloração em um dos cromossomos do par 7. Com a presunção de que todos os gametas desse homem são igualmente viáveis, qual proporção de seus filhos apresentará o mesmo cariótipo que ele apresenta?

22. Suponha que a meiose ocorre no estágio diploide temporário do ciclo de um organismo haploide de número de cromossomos *n*. Qual é a probabilidade de que uma célula haploide que resulta da divisão meiótica apresente um conjunto parental completo de centrômeros (ou seja, um conjunto todo de um genitor, ou todo do outro genitor)?

23. Finja que é o ano de 1868. Você é um fabricante de lentes jovem e habilidoso que trabalha em Viena. Com as suas novas lentes superiores, você acabou de construir um microscópio que apresenta melhor resolução do que quaisquer outros disponíveis. Em seu teste desse microscópio, você tem observado as células nos testículos de gafanhotos e está fascinado pelo comportamento das estranhas estruturas alongadas que observou dentro das células em divisão. Um dia, na biblioteca, você lê um artigo de periódico recente de G. Mendel sobre os "fatores" hipotéticos que ele alega explicarem os resultados de determinados cruzamentos em ervilhas. Em uma revelação súbita, você percebe os paralelos entre os seus estudos do gafanhoto e os estudos da ervilha de Mendel, e você resolve escrever uma carta para ele. O que você escreve? (O problema 23 tem por base uma ideia de Ernest Kroeker.)

24. A partir de um presumido cruzamento-teste $A/a \times a/a$, no qual *A* representa vermelho e *a* representa branco, utilize o teste do χ^2 para descobrir qual dos possíveis resultados a seguir corresponderia às expectativas:

a. 120 vermelhos, 100 brancos.

b. 5.000 vermelhos, 5.400 brancos.

c. 500 vermelhos, 540 brancos.

d. 50 vermelhos, 54 brancos.

25. Observe o quadrado de Punnett na Figura 3.4.

a. Quantos genótipos diferentes estão demonstrados nos 16 quadrados da grade?

b. Qual é a proporção genotípica subjacente à proporção fenotípica de 9:3:3:1?

c. Você consegue planejar uma fórmula simples para o cálculo da

quantidade de genótipos na progênie em cruzamentos di-híbridos, tri-híbridos e assim por diante? Repita em relação aos fenótipos.

d. Mendel previu que, dentro de todas as classes fenotípicas no quadrado de Punnett, com exceção de uma, deve haver diversos genótipos diferentes. Em particular, ele realizou muitos cruzamentos para identificar os genótipos subjacentes ao fenótipo amarelo e liso. Demonstre dois modos diferentes que poderiam ser utilizados para identificar os diversos genótipos subjacentes ao fenótipo amarelo e liso. (Relembre que todas as ervilhas amarelas e lisas aparentam ser idênticas.)

26. Presumindo a distribuição independente de todos os genes, desenvolva fórmulas que demonstrem o número de classes fenotípicas e o número de classes genotípicas a partir do autocruzamento de uma planta heterozigota para *n* pares de genes.

27. Nota: a primeira parte deste problema foi introduzida no Capítulo 2. A base lógica é estendida aqui.

Na planta *Arabidopsis thaliana*, um geneticista está interessado no desenvolvimento de tricomas (pequenas projeções) nas folhas. Uma grande triagem demonstra duas plantas mutantes (A e B) que não apresentam tricomas e essas mutantes aparentam ser possivelmente úteis no estudo do desenvolvimento dos tricomas. (Se eles forem determinados por mutações monogênicas, então o achado da função normal e anormal desses genes será instrutiva.) Cada planta foi cruzada com o tipo selvagem; em ambos os casos, a próxima geração (F_1) apresentou tricomas normais. Ao serem autofecundadas, as plantas da F_1 resultaram na seguinte F_2:

F_2 do mutante A: 602 normais; 198 sem tricomas.

F_2 do mutante B: 267 normais; 93 sem tricomas.

a. O que esses resultados demonstram? Inclua os genótipos propostos de todas as plantas em sua resposta.

b. Presuma que os genes estão localizados em cromossomos separados. Uma F_1 é produzida por meio do cruzamento da mutante A original com a mutante B original. Realiza-se o cruzamento-teste dessa F_1: que proporção

da progênie do cruzamento-teste não apresentará tricomas?

28. Em cães, a pelagem escura é dominante sobre a albina, e os pelos curtos são dominantes sobre os pelos longos. Presuma que esses efeitos sejam causados por dois genes de segregação independente. Foram realizados sete cruzamentos, conforme demonstrado a seguir, nos quais D e A fazem referência aos fenótipos escuro e albino, respectivamente, e S e L fazem referência aos fenótipos de pelos curtos e pelos longos.

Fenótipos parentais	*Número de progênie*			
	D, S	D, L	A, S	A, L
a. D, S × D, S	88	31	29	12
b. D, S × D, L	19	18	0	0
c. D, S × A, S	21	0	20	0
d. A, S × A, S	0	0	29	9
e. D, L × D, L	0	31	0	11
f. D, S × D, S	45	16	0	0
g. D, S × D, L	31	30	10	10

Escreva os genótipos dos progenitores em cada cruzamento. Utilize os símbolos *C* e *c* para os alelos da cor da pelagem escura e albina, e os símbolos *H* e *h* para os alelos dos pelos curtos e pelos longos, respectivamente. Presuma que os genitores são homozigotos, exceto se houver evidências em contrário.

29. Em tomates, um gene determina se a planta apresenta caules roxos (*G*) ou verdes (*g*), e um gene separado e independente determina se as folhas são *cut* (*P*) ou *potato* (*p*). Cinco cruzamentos de fenótipos da planta do tomate originam os resultados a seguir:

Cruzamento	Fenótipos parentais	*Número de progênie*			
		G, P	*G, p*	*g, P*	*g, p*
1	*G, P* × *g, P*	323	102	309	106
2	*G, P* × *G, p*	220	206	65	72
3	*G, P* × *g, P*	723	229	0	0
4	*G, P* × *g, p*	405	0	389	0
5	*G, p* × *g, P*	71	90	85	78

a. Quais alelos são dominantes?

b. Quais são os genótipos mais prováveis em relação aos genitores em cada

cruzamento?

30. Um alelo mutante em camundongos causa uma cauda dobrada. Seis pares de camundongos foram cruzados. Seus fenótipos e aqueles de sua progênie são fornecidos na tabela a seguir. N é o fenótipo normal; D é o fenótipo dobrado. Deduza modo de herança da cauda dobrada.

Cruzamento	*Genitores*		*Progênie*	
	♀	♂	♀	♂
1	N	D	Todos D	Todos N
2	D	N	$\frac{1}{2}$ D, $\frac{1}{2}$ N	$\frac{1}{2}$ D, $\frac{1}{2}$ N
3	D	N	Todos D	Todos D
4	N	N	Todos N	Todos N
5	D	D	Todos D	Todos D
6	D	D	Todos D	$\frac{1}{2}$ D, $\frac{1}{2}$ N

a. Ela é recessiva ou dominante?

b. Ela é autossômica ou ligada ao sexo?

c. Quais são os genótipos de todos os genitores e de todos os descendentes?

31. A cor normal dos olhos da *Drosophila* é vermelha, mas estão disponíveis linhagens nas quais todas as moscas apresentam olhos marrons. De modo semelhante, as asas normalmente são longas, mas existem cepas com asas curtas. Uma fêmea de uma linhagem pura com olhos marrons e asas curtas é cruzada com um macho de uma linhagem pura normal. A F_1 é composta por fêmeas normais e machos com asas curtas. Em seguida uma F_2 é produzida por meio do intercruzamento da F_1. *Ambos* os sexos das moscas da F_2

demonstram fenótipos como segue:

$\frac{3}{8}$ olhos vermelhos e asas longas
$\frac{3}{8}$ olhos vermelhos e asas curtas
$\frac{3}{8}$ olhos marrons e asas longas
$\frac{1}{8}$ olhos marrons e asas curtas

Deduza a herança desses fenótipos; utilize símbolos genéticos claramente definidos de sua própria invenção. Declare os genótipos de todas as três gerações e as proporções genotípicas da F_1 e da F_2.

Como solucionar o Problema 31

Antes de tentar uma solução para este problema, tente responder às questões a seguir:

1. O que a palavra *normal* significa neste problema?
2. As palavras *linhagem* e *cepa* são utilizadas neste problema. O que elas significam, e elas são intercambiáveis?
3. Desenhe um esboço simples das duas moscas parentais, demonstrando seus olhos, suas asas e as diferenças sexuais.
4. Quantas características diferentes existem neste problema?
5. Quantos fenótipos existem neste problema, e quais fenótipos estão associados a que características?
6. Qual é o fenótipo integral das fêmeas da F_1 denominadas “normais”?
7. Qual é o fenótipo integral dos machos da F_1 denominados “com asas curtas”?
8. Liste as proporções fenotípicas da F_2 em relação a cada caractere que você inventou na resposta à questão 4.
9. O que as proporções fenotípicas da F_2 informam a você?
10. Qual padrão importante de herança distingue a herança ligada ao sexo da herança autossômica?
11. Os dados da F_2 demonstram tal critério de distinção?
12. Os dados da F_1 demonstram um referido critério de distinção?

13. O que você pode aprender a respeito da dominância na F_1? Na F_2?
14. Quais regras a respeito do simbolismo do tipo selvagem você pode utilizar para decidir quais símbolos alélicos devem ser inventados em relação a esses cruzamentos?
15. O que significa “deduza a herança desses fenótipos”?

Agora tente solucionar o problema. Se você não conseguir fazer isso, crie uma lista de questões a respeito das coisas que você não compreende. Inspecione os objetivos do aprendizado no início do capítulo e se pergunte quais são relevantes para as suas questões. Se essa abordagem não funcionar, inspecione os Conceitos-chave deste capítulo e se pergunte quais poderiam ser relevantes para as suas questões.

32. Em uma população natural de plantas anuais, é encontrada uma única planta que aparenta estar enferma e que apresenta folhas amarelas. A planta é colhida e trazida para o laboratório. Observa-se que as taxas de fotossíntese são muito baixas. O pólen de uma planta normal com folhas verde-escuras é utilizado para fertilizar as folhas emasculadas da planta amarelada. Uma centena de sementes resulta, das quais apenas 60 germinam. Todas as plantas resultantes são de aspecto amarelo e enfermo.

a. Proponha uma explicação genética para o padrão de herança.

b. Sugira um teste simples para o seu modelo.

c. Explique a redução da fotossíntese, a enfermidade e o aspecto amarelado.

33. Qual é a base da variegação verde e branca nas folhas da *Mirabilis*? Se o cruzamento a seguir for realizado:

♀ variegada × ♂ verde

quais tipos de progênie podem ser previstos? O que dizer sobre o cruzamento recíproco?

34. Em *Neurospora,* o mutante *stp* exibe crescimento errático e intermitente.

Sabidamente o sítio mutante está no mtDNA. Se uma linhagem *stp* for utilizada como genitora em um cruzamento com uma linhagem normal que atua como masculina, qual tipo de progênie pode se esperado? O que dizer a respeito da progênie do cruzamento recíproco?

35. Duas plantas do milho são estudadas. Uma é resistente (R) e a outra é suscetível (S) a um determinado fungo patogênico. São realizados os cruzamentos a seguir, com os resultados demonstrados:

♀ S × ♂ R → Toda a progênie S
♀ R × ♂ S → Toda a progênie R

O que você pode concluir a respeito da localização dos determinantes genéticos de R e S?

36. Um di-híbrido presumido na *Drosophila*, *B*/*b*; *F*/*f*, é submetido ao cruzamento-teste com *b*/*b*; *f*/*f*. (*B* = Corpo preto; *b* = Corpo marrom; *F* = Cerdas bifurcadas; *f* = Cerdas não bifurcadas). Os resultados são:

Preto e bifurcado	230	Marrom e bifurcado	240
Preto e não bifurcado	210	Marrom e não bifurcado	250

Utilize o teste do χ^2 para determinar se esses resultados correspondem aos resultados esperados do cruzamento-teste do suposto di-híbrido.

37. Os números de progênie a seguir são consistentes com os resultados esperados do autocruzamento de uma planta que presumidamente é uma di-híbrida de dois genes de distribuição independente, *H*/*h*; *R*/*r*? (*H* = Folhas pilosas; *h* = Folhas lisas; *R* = Ovário redondo; *r* = Ovário alongado.) Explique a sua resposta.

Pilosa e redondo	178	Lisa e redondo	56
Pilosa e alongado	62	Lisa e alongado	24

38. Uma mariposa fêmea escura é cruzada com um macho escuro. Toda a

progênie masculina é escura, mas metade da progênie feminina é clara e o restante é escura. Proponha uma explicação para esse padrão de herança.

39. Em *Neurospora*, uma linhagem mutante denominada *stopper* (*stp*) surgiu espontaneamente. *Stopper* demonstrou crescimento errático que "para e começa", em comparação ao crescimento ininterrupto das linhagens do tipo selvagem. Nos cruzamentos, foram observados os resultados a seguir:

♀ *stopper* × ♂ do tipo selvagem → Toda a progênie *stopper*
♀ do tipo selvagem × ♂ *stopper* → Toda a progênie do tipo selvagem

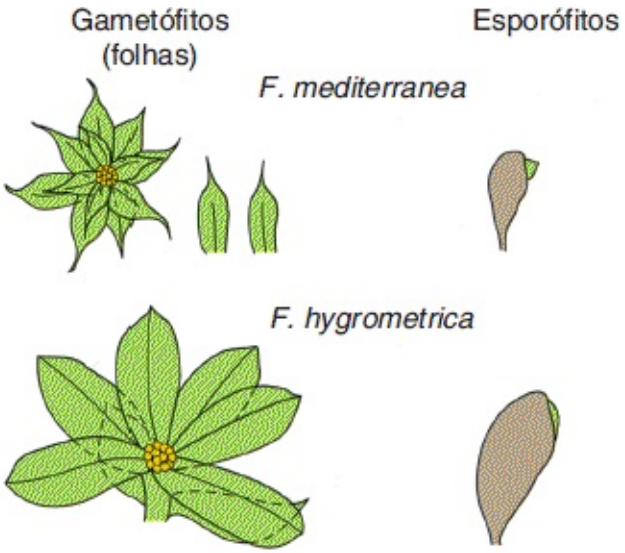
a. O que esses resultados sugerem a respeito da localização da mutação *stopper* no genoma?
b. De acordo com o seu modelo em relação à parte *a*, que progênie e proporções são previstas em óctades do cruzamento a seguir, incluindo uma mutação *nic3* localizada no cromossomo VI?

♀ *stp* · *nic3* × ♂ do tipo selvagem

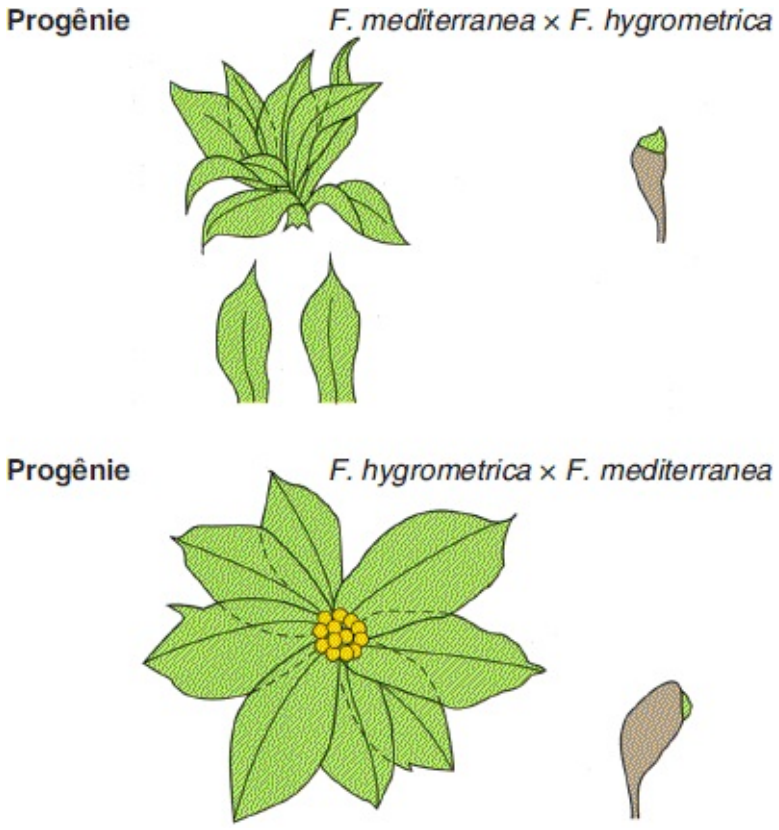
40. Em sistemas poligênicos, quantas classes fenotípicas correspondentes ao número de "doses" de poligenes são esperadas em autocruzamentos:
a. De linhagens com quatro poligenes heterozigotos?
b. De linhagens com seis poligenes heterozigotos?

41. Na autofecundação de um poligênico tri-híbrido R_1/r_1; R_2/r_2; R_3/r_3, utilize as regras do produto e da soma para calcular a proporção da progênie com apenas uma "dose" de poligenes.

42. Foram realizados cruzamentos recíprocos e autocruzamentos entre as duas espécies de musgos *Funaria mediterranea* e *F. hygrometrica*. Os esporófitos e as folhas dos gametófitos estão demonstrados no diagrama que acompanha:



Os cruzamentos estão escritos primeiramente com a genitora.



a. Descreva os resultados apresentados, resumindo os principais achados.

b. Proponha uma explicação dos resultados.

c. Demonstre como você testaria a sua explicação; assegure-se de

demonstrar como ela poderia ser distinguida de outras explicações.

43. Presuma que a planta diploide A apresenta um citoplasma geneticamente diferente daquele da planta B. Para estudar as relações nucleares e citoplasmáticas, você deseja obter uma planta com o citoplasma da planta A e o genoma nuclear predominantemente da planta B. Como você faria para produzir uma referida planta?

44. Você está estudando uma planta com um tecido composto por setores verdes e brancos. Você deseja decidir se esse fenômeno ocorre em virtude de (1) uma mutação de cloroplasto do tipo considerado neste capítulo, ou (2) uma mutação nuclear dominante que inibe a produção de clorofila e que está presente apenas em determinadas camadas de tecido da planta, como um mosaico. Resuma a abordagem experimental que você utilizaria para resolver este problema.

PROBLEMAS DESAFIADORES

45. Você tem três potes contendo bolas de gude, como segue:

pote 1	600 vermelhas	e	400 brancas
pote 2	900 azuis	e	100 brancas
pote 3	10 verdes	e	990 brancas

a. Se você selecionar às cegas uma bola de gude de cada pote, calcule a probabilidade de obter:

(1) uma vermelha, uma azul e uma verde.
(2) três brancas.
(3) uma vermelha, uma verde e uma branca.
(4) uma vermelha e duas brancas.
(5) uma colorida e duas brancas.
(6) no mínimo uma branca.

b. Em uma determinada planta, *R* = Vermelho e *r* = Branco. Você realiza a autofecundação de uma heterozigota *R*/*r* vermelha com a finalidade expressa de obter uma planta branca para um experimento. Qual número

mínimo e sementes você deve cultivar para ter no mínimo 95% de certeza de obter no mínimo um indivíduo branco?

c. Quando uma mulher recebe a injeção de um ovócito fertilizado *in vitro*, a probabilidade de sucesso na sua implantação é de 20%. Se uma mulher receber a injeção de cinco ovócitos fertilizados simultaneamente, qual é a probabilidade de que ela engravidará? (A parte *c* é de Margaret Holm.)

46. Em tomates, o fruto vermelho é dominante sobre o amarelo, o fruto biloculado é dominante sobre o fruto multiloculado, e o pé alto é dominante sobre o anão. Um cultivador possui duas linhagens puras: (1) vermelha, biloculada e anã, e (2) amarela, multiloculada e alta. A partir dessas duas linhagens, ele deseja produzir uma nova linhagem pura para comercialização que seja amarela, biloculada e alta. Como exatamente ele deve fazer isso? Demonstre não apenas quais cruzamentos devem ser realizados, mas também quantos descendentes devem ser amostrados em cada caso.

47. Lidamos principalmente com apenas dois genes, mas os mesmos princípios se mantêm em relação a mais de dois genes. Considere o cruzamento a seguir:

A/*a*; *B*/*b*; *C*/*c*; *D*/*d*; *E*/*e* × *a*/*a*; *B*/*b*; *c*/*c*; *D*/*d*; *e*/*e*

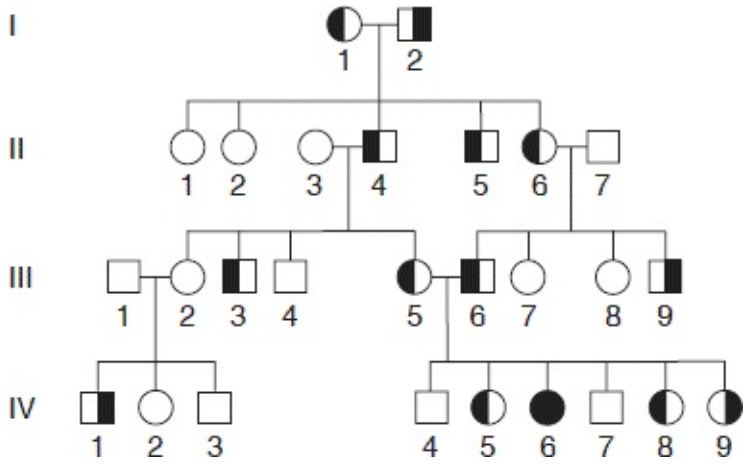
a. Qual proporção da progênie *fenotipicamente* se assemelhará (1) ao primeiro genitor, (2) ao segundo genitor, (3) a qualquer genitor e (4) a nenhum dos genitores?

b. Qual proporção da progênie será *genotipicamente* a mesma (1) do primeiro genitor, (2) do segundo genitor, (3) de qualquer genitor e (4) de nenhum genitor?

Presuma a segregação independente.

48. O heredograma a seguir demonstra o padrão da transmissão de dois fenótipos humanos raros: catarata e nanismo pituitário. Os familiares com catarata estão demonstrados com uma metade esquerda sólida do símbolo; aqueles com nanismo pituitário estão indicados com uma metade direita

sólida.



a. Qual é o modo mais provável de herança de cada um desses fenótipos? Explique.

b. Liste os genótipos de todos os membros na geração III, tanto quanto possível.

c. Se um cruzamento hipotético ocorreu entre IV-1 e IV-5, qual é a probabilidade de que o primeiro filho seja um anão com catarata? Uma criança fenotipicamente normal? (O Problema 48 é adaptado de J. Kuspira e R. Bhambhani, *Compendium of Problems in Genetics*. Direitos autorais 1994 por Wm. C. Brown.)

49. Um geneticista de milho possui três linhagens puras de genótipos *a*/*a*; *B*/*B*; *C*/*C*, *A*/*A*; *b*/*b*; *C*/*C* e *A*/*A*; *B*/*B*; *c*/*c*. Todos os fenótipos determinados por *a*, *b* e *c* aumentarão o valor de mercado do milho. Assim, naturalmente ele deseja combinar todos em uma linhagem pura do genótipo *a*/*a*; *b*/*b*; *c*/*c*.

a. Resuma um programa de cruzamento efetivo que possa ser utilizado para obter a linhagem pura *a*/*a*; *b*/*b*; *c*/*c*.

b. A cada estágio, declare exatamente quais fenótipos serão selecionados e forneça as suas frequências esperadas.

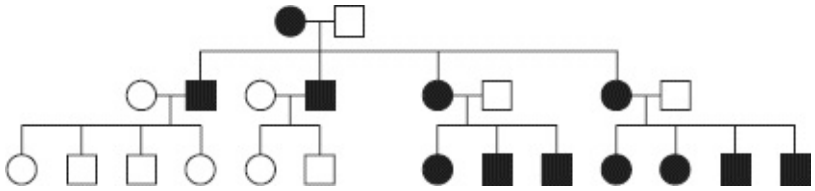
c. Existe mais de um modo de obter o genótipo desejado? Qual é o melhor modo?

Presuma a distribuição independente dos três pares de genes. (Nota: o

milho será submetido facilmente à autopolinização ou à polinização cruzada.)

50. Em humanos, a visão colorida depende de genes que codificam três pigmentos. Os genes *R* (pigmento vermelho) e *G* (pigmento verde) estão próximos no cromossomo X, enquanto o gene *B* (pigmento azul) é autossômico. Uma mutação recessiva em qualquer um desses genes pode causar daltonismo. Suponha que um homem daltônico se casou com uma mulher com visão para cores normal. Os quatro filhos desse casamento eram daltônicos e as cinco filhas eram normais. Especifique os genótipos mais prováveis de ambos os genitores e de seus filhos, explicando a sua justificativa. (Um desenho do heredograma provavelmente será útil.) (O Problema 50 é de Rosemary Redfield.)

51. Considere o heredograma a seguir em relação a uma doença muscular humana rara.



a. Qual característica incomum distingue esse heredograma daqueles estudados anteriormente neste capítulo?

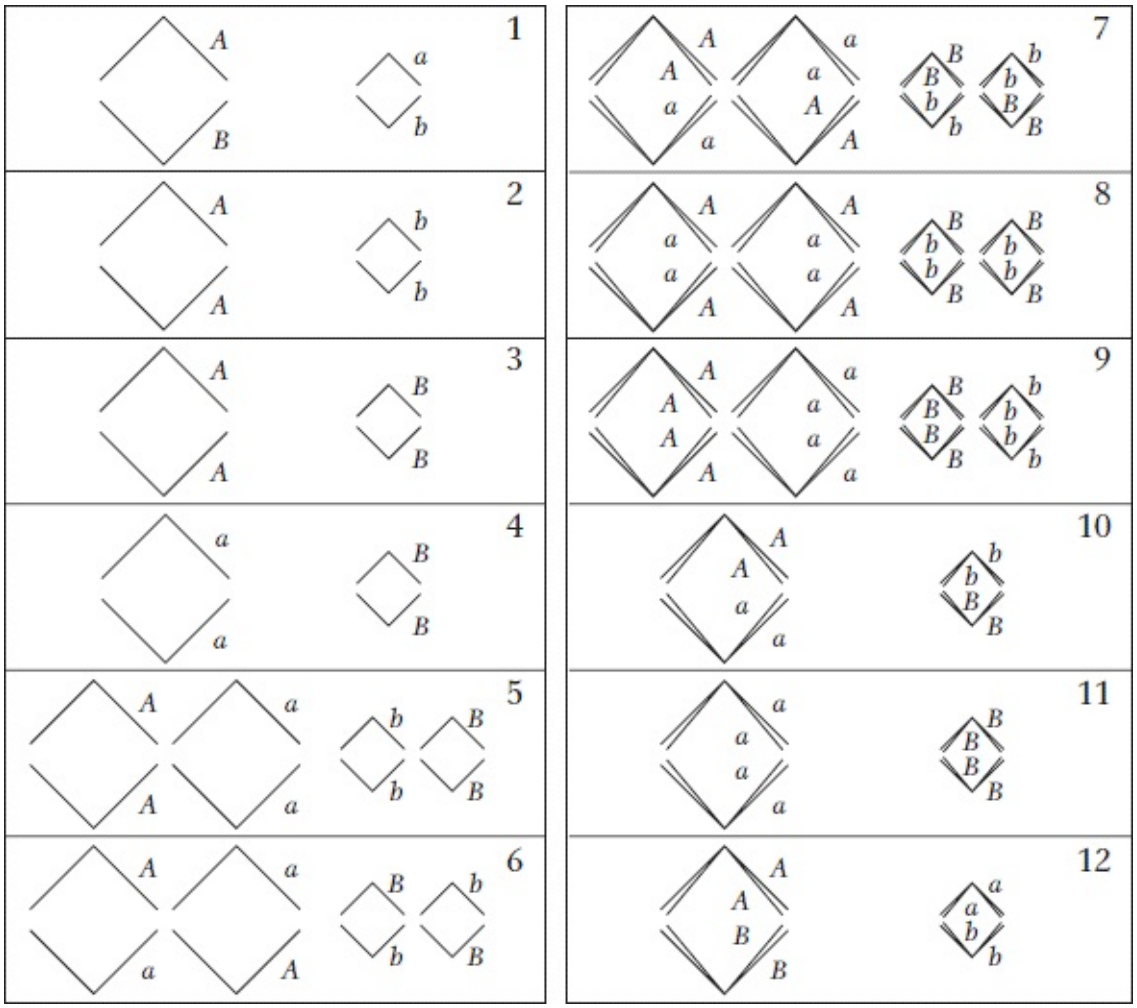
b. Onde você acredita que o DNA mutante responsável por esse fenótipo está localizado na célula?

52. A planta *Haplopappus gracilis* apresenta 2*n* de 4. Foi estabelecida uma cultura de célula diploide e, na fase S pré-mitótica, um nucleotídio radioativo foi adicionado e incorporado ao DNA recentemente sintetizado. As células então foram removidas da radioatividade, lavadas e possibilitou-se que prosseguissem pela mitose. Os cromossomos, ou cromátides, radioativos podem ser detectados por meio da colocação de emulsão fotográfica sobre as células; os cromossomos radioativos apareceram cobertos por manchas de prata da emulsão. (Os cromossomos

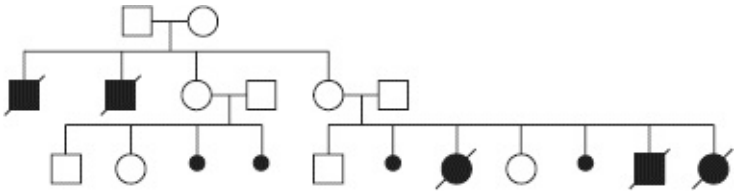
"tiram a sua própria fotografia".) Desenhe os cromossomos na prófase e na telófase da primeira e da segunda divisões da mitose após o tratamento radioativo. Se eles forem radioativos, desenhe no seu diagrama. Se houver diversas possibilidades, demonstre-as também.

53. Na espécie do Problema 52, você pode introduzir a radioatividade por meio da injeção nas anteras na fase S antes da meiose. Desenhe os quatro produtos da meiose com seus cromossomos e demonstre quais são radioativos.

54. A planta *Haplopappus gracilis* é diploide e $2n = 4$. Existe um par longo e um par curto de cromossomos. Os diagramas a seguir (numerados de 1 a 12) representam as anáfases (estágios de "separação") de células individuais na meiose ou na mitose em uma planta que é geneticamente di-híbrida (*A*/*a*; *B*/*b*) em relação a genes em cromossomos diferentes. As linhagens representam cromossomos ou cromátides, e as pontas dos V representam os centrômeros. Em cada caso, indique se o diagrama representa uma célula em meiose I, meiose II ou mitose. Se um diagrama demonstrar uma situação impossível, afirme isso.



55. O heredograma a seguir demonstra a recorrência de uma doença neurológica rara (símbolos pretos grandes) e abortos fetais espontâneos (símbolos pretos pequenos) em uma família. (Uma barra diagonal significa que o indivíduo é falecido.) Forneça uma explicação para este heredograma em relação à segregação citoplasmática de mitocôndrias defeituosas.



56. Um homem é braquidáctilo (dedos das mãos muito curtos; autossômico dominante raro) e sua esposa não é. Ambos conseguem degustar a

substância química feniltiocarbamida (autossômico dominante; alelo comum), mas suas mães não conseguiam.

a. Forneça os genótipos do casal.

Se os genes se distribuem independentemente e o casal tem quatro filhos, qual é a probabilidade de:

b. Todos eles serem braquidáctilos?

c. Nenhum ser braquidáctilo?

d. Todos eles serem degustadores?

e. Todos eles serem não degustadores?

f. Todos eles serem degustadores braquidáctilos?

g. Nenhum ser degustador braquidáctilo?

h. No mínimo um ser um degustador braquidáctilo?

57. Uma forma de esterilidade masculina no milho é transmitida maternalmente. Plantas de uma linhagem masculina estéril cruzadas com pólen normal originam plantas masculinas estéreis. Além disso, algumas linhagens de milho sabidamente carreiam um alelo restaurador nuclear dominante (*Rf*) que restaura a fertilidade do pólen em linhagens masculinas estéreis.

a. Pesquisas demonstram que a introdução de alelos restauradores em linhagens masculinas estéreis não altera nem afeta a manutenção dos fatores citoplasmáticos em relação à esterilidade masculina. Qual tipo de resultados de pesquisas levaria a uma referida conclusão?

b. Uma planta masculina estéril é cruzada com o pólen de uma planta homozigota em relação a *Rf*. Qual é o genótipo da F_1? O fenótipo?

c. As plantas da F_1 da parte *b* são utilizadas como fêmeas em um cruzamento-teste com o pólen de uma planta normal (*rf*/*rf*). Quais são os resultados desse cruzamento-teste? Forneça os genótipos e os fenótipos, e designe o tipo de citoplasma.

d. O alelo restaurador já descrito pode ser denominado *Rf-1*. Outro restaurador dominante, *Rf-2*, foi observado. *Rf-1* e *Rf-2* estão localizados em cromossomos diferentes. Cada um ou ambos os alelos restauradores proporcionarão a fertilidade ao pólen. Com a utilização de uma planta

masculina estéril como uma testadora, qual será o resultado de um cruzamento no qual o progenitor do sexo masculino é:

(1) heterozigoto em ambos os *loci* restauradores?

(2) homozigoto dominante em um *locus* restaurador e homozigoto recessivo no outro?

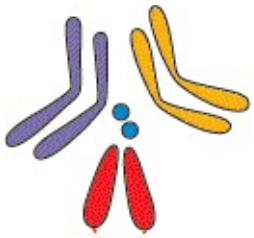
(3) heterozigoto em um *locus* restaurador e homozigoto recessivo no outro?

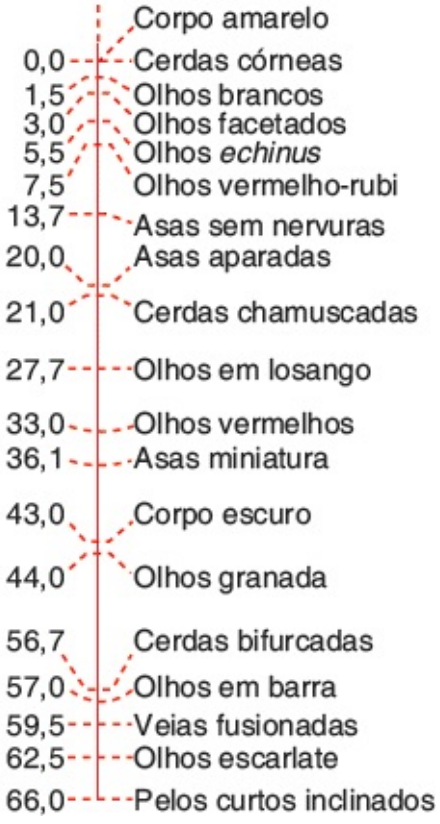
(4) heterozigoto em um *locus* restaurador e homozigoto dominante no outro?

CAPÍTULO

4

Mapeamento de Cromossomos Eucarióticos por Recombinação





Mapa com base na recombinação de um dos cromossomos de *Drosophila* (*organismo na imagem*), demonstrando os *loci* dos genes cujas mutações produzem fenótipos conhecidos. (©David Scharf/Corbis.)

TÓPICOS

4.1 Diagnóstico de ligação
4.2 Mapeamento por frequência de recombinantes
4.3 Mapeamento com marcadores moleculares
4.4 Mapeamento de centrômeros com tétrades lineares
4.5 Utilização do teste do qui-quadrado para inferir ligação
4.6 Cômputo de crossovers múltiplos não visualizados
4.7 Utilização de mapas com base em recombinação em conjunto com mapas físicos
4.8 Mecanismo molecular de crossover

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Realizar uma análise quantitativa da progênie de um cruzamento-teste di-híbrido para avaliar se os dois genes estão ligados no mesmo cromossomo ou não
- Estender o mesmo tipo de análise a diversos *loci* para produzir um mapa

das posições relativas dos *loci* em um cromossomo
- Em fungos ascomicetos, mapear os centrômeros a outros *loci* ligados
- Em ascos, prever as proporções de alelos oriundos de etapas específicas no modelo heterodúplex de *crossing over*.

Algumas das questões a que os geneticistas desejam responder a respeito do genoma são: "*Que genes* estão presentes no genoma?"; "*Que funções* apresentam?"; "*Que posições* ocupam nos cromossomos?" Convencionou-se chamar "mapeamento" a busca pela resposta à terceira questão. O mapeamento é o principal enfoque deste capítulo, mas todas as três questões estão inter-relacionadas, conforme veremos.

Sentimos diariamente quão importantes são os mapas em geral e é fato que, em algum momento de nossas vidas, já os utilizamos para encontrar o caminho. Relevante para o enfoque deste capítulo é que, em algumas situações, diversos mapas precisam ser utilizados simultaneamente. Um bom exemplo na vida diária está em percorrer o denso arranjo de ruas e edifícios de uma cidade tal como Londres, na Inglaterra. Um mapa das ruas que demonstra o plano geral é uma necessidade. Entretanto, o mapa das ruas é utilizado por turistas e também por londrinos juntamente com outro mapa, aquele do sistema ferroviário subterrâneo. O sistema subterrâneo é tão complexo e intrincado que, em 1933, um engenheiro de circuitos elétricos chamado Harry Beck desenhou o mapa simplificado (embora distorcido) que permanece até hoje um ícone de Londres. Os mapas das ruas e dos subterrâneos de Londres estão comparados na Figura 4.1. Observe que as posições das estações subterrâneas e as distâncias exatas entre elas não são de interesse por si próprias, exceto como um modo para se chegar até um destino de interesse, tal como a Abadia de Westminster. Veremos três paralelos com os mapas de Londres quando os mapas cromossômicos são utilizados para indicar "destinos" individuais ou genes específicos. Primeiramente, diversos tipos de mapas cromossômicos com frequência são necessários e precisam ser utilizados em conjunto; em segundo lugar, mesmo mapas que contêm distorções são úteis; e, em terceiro lugar, diversos sítios em um mapa cromossômico são traçados apenas

por serem úteis para tentar situar outros pontos que são os de real interesse.

A obtenção de um mapa das posições dos genes nos cromossomos é um esforço que ocupou milhares de geneticistas durante os últimos 80 anos. Por que isso é tão importante? Existem diversos motivos:

1. A posição do gene é uma informação crucial para construir *genótipos complexos* necessários para fins experimentais ou para aplicações comerciais. Por exemplo, no Capítulo 6, veremos casos nos quais combinações alélicas especiais precisam ser reunidas para explorar a interação gênica.
2. O conhecimento sobre a posição ocupada por um gene proporciona um modo para a descoberta da sua *estrutura e função*. A posição de um gene pode ser utilizada para defini-lo no nível do DNA. Por sua vez, a sequência de DNA de um gene do tipo selvagem ou de seu alelo mutante é uma parte necessária para a dedução da sua função subjacente.
3. Os genes presentes e seu arranjo nos cromossomos com frequência são discretamente diferentes em espécies correlatas. Por exemplo, o longo cromossomo humano número 2 é dividido em dois cromossomos mais curtos nos grandes primatas. Ao comparar essas diferenças, geneticistas podem deduzir os *mecanismos genéticos evolutivos* por meio dos quais esses genomas divergiram. Portanto, mapas cromossômicos são úteis na interpretação dos mecanismos da evolução.

FIGURA 4.1 Estes mapas de Londres ilustram o princípio de que, com frequência, diversos mapas são necessários para se chegar ao destino desejado. O mapa da ferrovia subterrânea ("o Tubo") é utilizado para se chegar até um destino de interesse, tal como o endereço de uma rua, demonstrado no mapa das ruas. Em genética, dois tipos diferentes de mapas genômicos com frequência são úteis para a localizar um gene, levando a uma compreensão sobre a sua estrutura e função. (*A.* ©MAPS.com/Corbis; *B.* Transport for London.)

O arranjo dos genes nos cromossomos é representado diagramaticamente como um **mapa cromossômico** unidimensional, que demonstra as posições dos genes, conhecidas como ***loci*** (no singular, *locus*), e as distâncias entre os *loci* com base em algum tipo de escala. Dois tipos básicos de mapas cromossômicos são utilizados atualmente em genética; eles são montados por meio de procedimentos consideravelmente diferentes, ainda que sejam utilizados de modo complementar. Os *mapas baseados na recombinação*, que são o tópico deste capítulo, mapeiam os *loci* dos genes que foram identificados por meio de fenótipos mutantes que demonstram herança monogênica. Os *mapas físicos* (ver Capítulo 14) demonstram os genes como segmentos arranjados ao longo da molécula de DNA que constitui um cromossomo. Esses mapas demonstram visões diferentes do genoma, mas, assim como os mapas de Londres, podem ser utilizados em conjunto para se descobrir a função de um gene no nível molecular e como essa função influencia o fenótipo.

CONCEITO-CHAVE Os mapas genéticos são úteis para a construção de linhagens, para a interpretação de mecanismos evolutivos e para a descoberta da função desconhecida de um gene, essa última facilitada pela integração das informações baseadas em recombinação com aquelas obtidas de mapas físicos.

4.1 Diagnóstico de ligação

Os **mapas de recombinação** dos cromossomos normalmente são montados com dois ou três genes por vez, com a utilização de um método denominado análise de ligação. Quando os geneticistas dizem que dois genes estão **ligados**, eles querem dizer que os *loci* daqueles genes estão no mesmo cromossomo e, portanto, os alelos em qualquer um dos homólogos estão fisicamente unidos (ligados) pelo DNA entre eles. O modo como os primeiros geneticistas deduziram a ligação é um meio útil para apresentar a maior parte das ideias-chave e dos procedimentos-chave na análise.

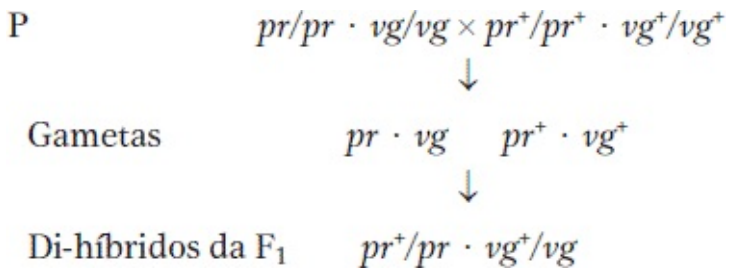
Utilização da frequência de recombinantes para reconhecer ligação

No início do século 20, William Bateson e R. C. Punnett (em homenagem ao qual o quadrado de Punnett foi denominado) estavam estudando a herança de dois genes em ervilhas-de-cheiro. Em uma autofecundação-padrão de uma F_1 di-híbrida, a F_2 não demonstrou a proporção de 9:3:3:1 prevista pelo princípio da distribuição independente. De fato, Bateson e Punnett observaram que determinadas combinações de alelos apareceram com mais frequência do que o esperado, quase como se estivessem fisicamente unidas de algum modo. Entretanto, eles não tinham uma explicação para essa descoberta.

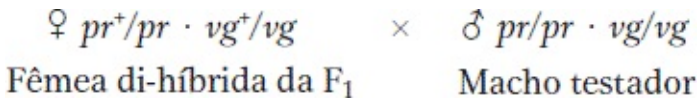
Mais tarde, Thomas Hunt Morgan observou um desvio semelhante da segunda lei de Mendel enquanto estudava dois genes autossômicos em *Drosophila*. Morgan propôs a ligação como uma hipótese para explicar o fenômeno da aparente associação de alelos.

Vejamos alguns dos dados de Morgan. Um dos genes afetava a cor dos olhos

(*pr*, roxo, e pr^+, vermelho) e o outro gene afetava o comprimento das asas (*vg*, vestigial, e vg^+, normal). (As asas vestigiais são muito pequenas em comparação ao tipo selvagem.) Os alelos do tipo selvagem de ambos os genes são dominantes. Morgan realizou um cruzamento para obter di-híbridos e em seguida prosseguiu com um cruzamento-teste:



Cruzamento-teste:



Fêmea di-híbrida da F_1 Macho testador

A utilização do cruzamento-teste por Morgan é importante. Tendo em vista que o genitor testador contribui com gametas que carreiam apenas alelos recessivos, os fenótipos da descendência revelam diretamente os alelos trazidos pelos gametas do genitor di-híbrido, conforme descrito nos Capítulos 2 e 3. Portanto, o analista pode se concentrar na meiose em um genitor (o di-híbrido) e essencialmente esquecer a meiose no outro (o testador). Contrariamente, a partir de uma F_1 *autofecundada*, existem dois conjuntos de meioses a serem considerados na análise da progênie: um no genitor e o outro na genitora.

Os resultados do cruzamento-teste de Morgan foram como segue (listados como classes gaméticas do di-híbrido):

$pr^+ \cdot vg^+$	1.339
$pr \cdot vg$	1.195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154

2.839

Obviamente, esses números desviam-se drasticamente da previsão mendeliana de uma proporção de 1:1:1:1 esperada a partir da distribuição independente (aproximadamente 710 em cada uma das quatro classes). Nos resultados de Morgan, observamos que as primeiras duas combinações de alelos estão, em sua maioria, indicando claramente que eles estão associados, ou "ligados".

Outro modo útil de avaliar os resultados do cruzamento-teste é por meio da consideração da porcentagem de recombinantes na progênie. Por definição, os recombinantes no presente cruzamento são de dois tipos, $pr^+ \cdot vg$ e $pr \cdot vg^+$, tendo em vista que eles claramente não são os dois genótipos fornecidos para o di-híbrido da F_1 pelas moscas parentais homozigotas originais (mais precisamente, por seus gametas). Observamos que os dois tipos recombinantes são aproximadamente iguais em frequência (151 ~ 154). Seu total é 305, que é uma frequência de (305/2.839) × 100 ou 10,7%. Podemos compreender esses dados, como Morgan compreendeu, ao postular que os genes estavam ligados no mesmo cromossomo e, assim, as combinações alélicas parentais são mantidas juntas na maior parte da progênie. No di-híbrido, a conformação alélica é necessariamente como segue:

$$\frac{pr^+ \quad vg^+}{pr \quad vg}$$

A tendência de alelos ligados serem herdados juntos está ilustrada na Figura 4.2.

Agora observemos outro cruzamento que Morgan realizou com a utilização dos mesmos alelos, mas em uma combinação diferente. Nesse cruzamento, cada genitor é homozigoto em relação ao alelo do tipo selvagem de um gene e ao alelo mutante do outro. Novamente, as fêmeas da F_1 foram submetidas ao cruzamento-teste:

P $pr^+/pr^+ \cdot vg/vg \times pr/pr \cdot vg^+/vg^+$

↓

Gametas $pr^+ \cdot vg$ $pr \cdot vg^+$

↓

Di-híbrido da F_1 $pr^+/pr \cdot vg^+/vg$

Cruzamento-teste:

♀ $pr^+/pr \cdot vg^+/vg$ × ♂ $pr/pr \cdot vg/vg$

Fêmea di-híbrida da F_1 Macho testador

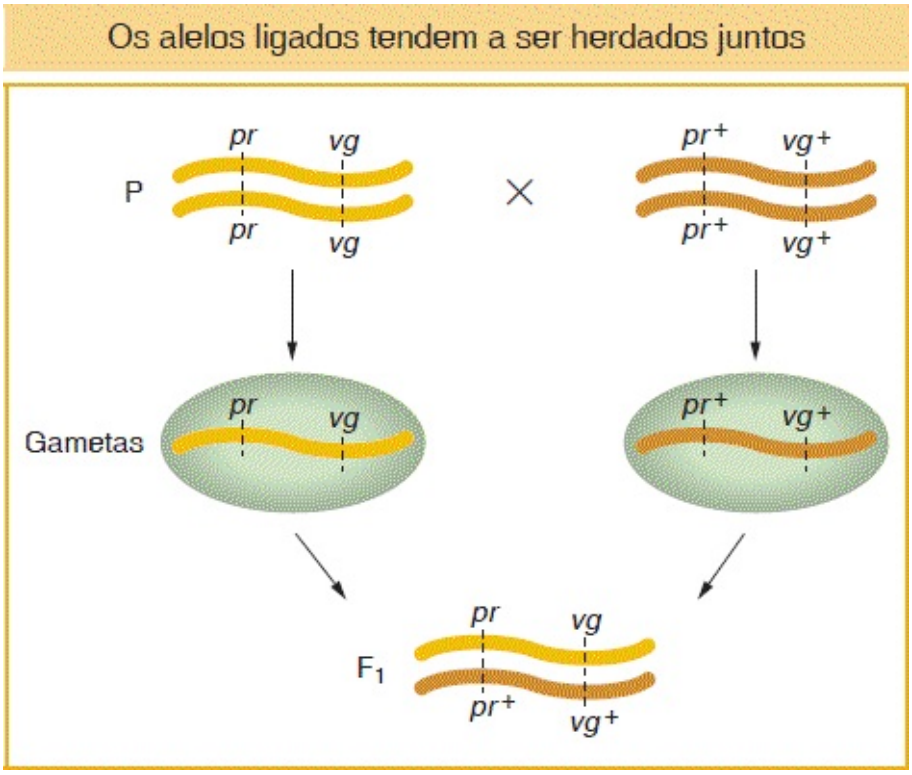


FIGURA 4.2 Herança simples de dois genes localizados no mesmo par de cromossomos. Os mesmos genes estão presentes juntos em um cromossomo em ambos os genitores e na progênie.

A progênie a seguir foi obtida a partir do cruzamento-teste:

$pr^+ \cdot vg^+$	157
$pr \cdot vg$	146
$pr^+ \cdot vg$	965

$pr \cdot vg^+$	1.067
	2.335

Novamente, esses resultados não estão nem sequer próximos de uma proporção mendeliana de 1:1:1:1. Entretanto, agora as classes recombinantes são o contrário daquelas na primeira análise, pr^+ vg^+ e *pr vg*. Entretanto, observe que a sua frequência é aproximadamente a mesma: (157 + 146)/2.335 × 100 = 12,9%. Novamente, é sugerida a ligação, mas, nesse caso, o di-híbrido da F_1 é necessariamente como segue:

$$\frac{pr^+ \quad vg}{pr \quad vg^+}$$

Resultados de cruzamentos-teste de di-híbridos como aqueles há pouco apresentados são comumente encontrados em genética. Eles seguem o padrão geral:

Duas classes de não recombinantes igualmente frequentes que totalizam *mais* de 50%
Duas classes de recombinantes igualmente frequentes que totalizam menos de 50%

> CONCEITO-CHAVE Quando dois genes estão próximos no mesmo par cromossômico (ou seja, estão ligados), eles não se distribuem independentemente, mas produzem uma frequência de recombinantes inferior a 50%. Portanto, uma frequência de recombinantes inferior a 50% é diagnóstica de ligação.

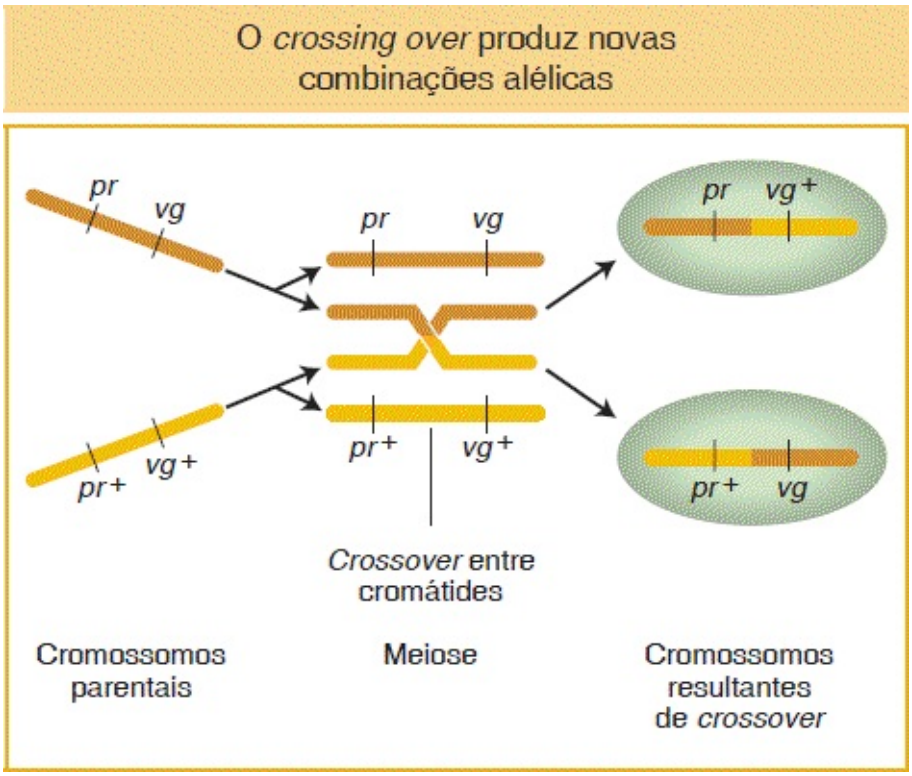


FIGURA 4.3 A troca de partes por meio do *crossing over* produz cromossomos gaméticos cujas combinações alélicas diferem das combinações parentais.

Como os crossovers produzem recombinantes de genes ligados

A hipótese de ligação explica por que as combinações alélicas das gerações parentais permanecem juntas: os genes estão fisicamente unidos pelo segmento do cromossomo entre eles. Mas exatamente como *quaisquer* recombinantes são produzidos quando os genes estão ligados? Morgan sugeriu que, quando cromossomos homólogos pareiam na meiose, os cromossomos ocasionalmente se rompem e trocam partes, em um processo denominado ***crossing over***. A Figura 4.3 ilustra essa troca física de segmentos cromossômicos. As duas novas combinações são denominadas **produtos de *crossover***.

Existe algum processo microscopicamente observável que possa explicar o *crossing over*? Na meiose, quando os cromossomos homólogos duplicados pareiam entre si — em termos genéticos, quando as duas díades se unem como um bivalente — com frequência ocorre a formação de uma estrutura com formato de cruz, denominada *quiasma*, entre duas cromátides não irmãs. Para Morgan, o

aparecimento dos quiasmas comprovou visualmente o conceito de *crossing over*. (Observe que os quiasmas parecem indicar que as *cromátides*, não os cromossomos duplicados, participam de um *crossover*. Retornaremos a esse ponto posteriormente.)

> CONCEITO-CHAVE Em relação aos genes ligados, os recombinantes são produzidos por meio de *crossovers*. Os quiasmas são as manifestações visíveis dos *crossovers*.

Simbolismo e terminologia de ligação

O trabalho de Morgan demonstrou que os genes ligados em um di-híbrido podem estar presentes em uma de duas conformações básicas. Em uma, os dois alelos dominantes, ou do tipo selvagem, estão presentes no mesmo homólogo (como na Figura 4.3); esse arranjo é denominado **conformação cis** (*cis* significa "adjacente"). Na outra, eles estão em homólogos diferentes, o que é denominado **conformação trans** (*trans* significa "oposto"). As duas conformações são escritas como segue:

Cis	*AB/ab* ou + +/*ab*
Trans	*Ab/aB* ou + *b/a* +

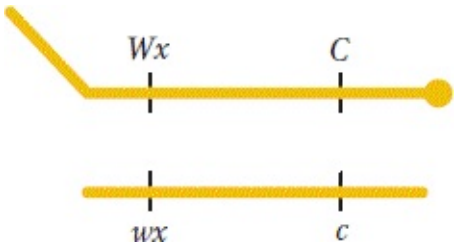
Observe as convenções a seguir que se referem ao simbolismo da ligação:

1. Alelos no mesmo homólogo não apresentam pontuação entre si.
2. Uma barra separa simbolicamente os dois homólogos.
3. Os alelos sempre são escritos na mesma ordem em cada homólogo.
4. Como nos capítulos anteriores, os genes sabidamente localizados em cromossomos diferentes (genes não ligados) são demonstrados em separado por um ponto e vírgula — por exemplo, *A/a; C/c*.
5. Neste livro, genes de ligação *desconhecida* são demonstrados em separado por um ponto, *A/a · D/d*.

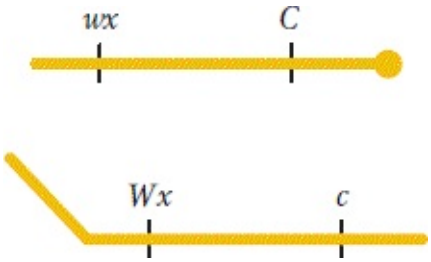
Evidências de que o crossing over é um processo de quebra e reunião

A ideia de que os recombinantes são produzidos por meio de algum tipo de troca de material entre cromossomos homólogos era uma ideia atraente. Mas era necessário experimentação para testar essa hipótese. Uma primeira etapa era encontrar um caso no qual a troca de partes entre os cromossomos fosse visível ao microscópio. Diversos investigadores abordaram esse problema do mesmo modo e uma de suas análises se segue.

Em 1931, Harriet Creighton e Barbara McClintock estavam estudando dois genes do milho que elas sabiam estarem ambos localizados no cromossomo 9. Um afetava a cor da semente (*C*, colorida; *c*, incolor) e o outro afetava a composição do endosperma (*Wx*, ceroso; *wx*, amiláceo). A planta era uma di-híbrida em conformação cis. Entretanto, em uma planta, o cromossomo 9 que carreava os alelos *C* e *Wx* era incomum, no sentido em que ele também carreava um elemento grande e densamente corado (denominado *knob*) na extremidade *C* e um segmento mais longo do cromossomo na extremidade *Wx*; portanto, o heterozigoto era:



Na progênie de um cruzamento-teste dessa planta, elas compararam os genótipos recombinantes e parentais. Elas observaram que todos os recombinantes herdaram um ou outro dos dois cromossomos a seguir, dependendo de sua constituição recombinante:



Portanto, havia uma correlação precisa entre o evento *genético* do aparecimento de recombinantes e o evento *cromossômico* do *crossing over*. Consequentemente, os quiasmas aparentavam ser os sítios de troca, embora o que foi considerado como teste definitivo só tenha sido realizado em 1978.

O que podemos dizer a respeito do mecanismo molecular da troca cromossômica em um evento de *crossover*? A resposta curta é que o *crossover* resulta da quebra e da reunião do DNA. Dois cromossomos parentais são quebrados na mesma posição e em seguida cada parte é unida à parte vizinha do *outro* cromossomo. Na Seção 4.8, veremos um modelo dos processos moleculares que possibilitam que o DNA seja quebrado e reunido de uma maneira precisa, de modo que não haja perda ou ganho de material genético.

> CONCEITO-CHAVE Um *crossover* é a quebra de duas moléculas de DNA na mesma posição e a sua reunião em duas combinações recombinantes recíprocas.

Evidências de que o crossing over ocorre no estágio de quatro cromátides

Conforme já observado, a representação diagramática do *crossing over* na Figura 4.3 demonstra a ocorrência de um *crossover* no estágio de quatro cromátides da meiose; em outras palavras, os *crossovers* ocorrem entre *cromátides* não irmãs. Entretanto, *teoricamente* era possível que o *crossing over* ocorresse antes da replicação, no estágio de *dois* cromossomos. Essa incerteza foi resolvida por meio da análise genética de organismos cujos quatro produtos da meiose permanecem juntos em grupos de quatro, denominados *tétrades*. Esses organismos, que encontramos nos Capítulos 2 e 3, são os fungos e as algas unicelulares. Os produtos da meiose de uma única tétrade podem ser isolados, o que é equivalente ao isolamento de todas as quatro cromátides de um único meiócito. As análises de tétrades de cruzamentos *nos quais os genes estão ligados* demonstram muitas tétrades que contêm quatro diferentes combinações

alélicas. Por exemplo, a partir do cruzamento:

AB × *ab*

algumas (mas não todas) tétrades contêm quatro genótipos:

AB
Ab
aB
ab

Esse resultado pode ser explicado apenas se *crossovers* ocorrerem no estágio de quatro cromátides, tendo em vista que, se *crossovers* ocorressem no estágio de dois cromossomos, somente poderia haver um máximo de dois genótipos diferentes em uma tétrade individual. Esse argumento está ilustrado na Figura 4.4.

Crossovers múltiplos podem incluir mais de duas cromátides

A análise de tétrades também pode demonstrar duas outras características importantes do *crossing over*. Primeiramente, em alguns meiócitos podem ocorrer diversos *crossovers* ao longo de um par de cromossomos. Em segundo lugar, em qualquer meiócito, esses *crossovers* múltiplos podem trocar o material entre mais de duas cromátides. Para pensar a respeito desse assunto, precisamos observar o caso mais simples: *crossovers* duplos. Para estudar os *crossovers* duplos, precisamos de três genes ligados. Por exemplo, se todos os três *loci* estiverem ligados em um cruzamento tal como:

ABC × *abc*

são possíveis muitos tipos de tétrades diferentes, mas alguns tipos são informativos na presente conexão, tendo em vista que eles podem ser explicados apenas por *crossovers* duplos nos quais participam mais de duas cromátides. Considere a tétrade a seguir como um exemplo:

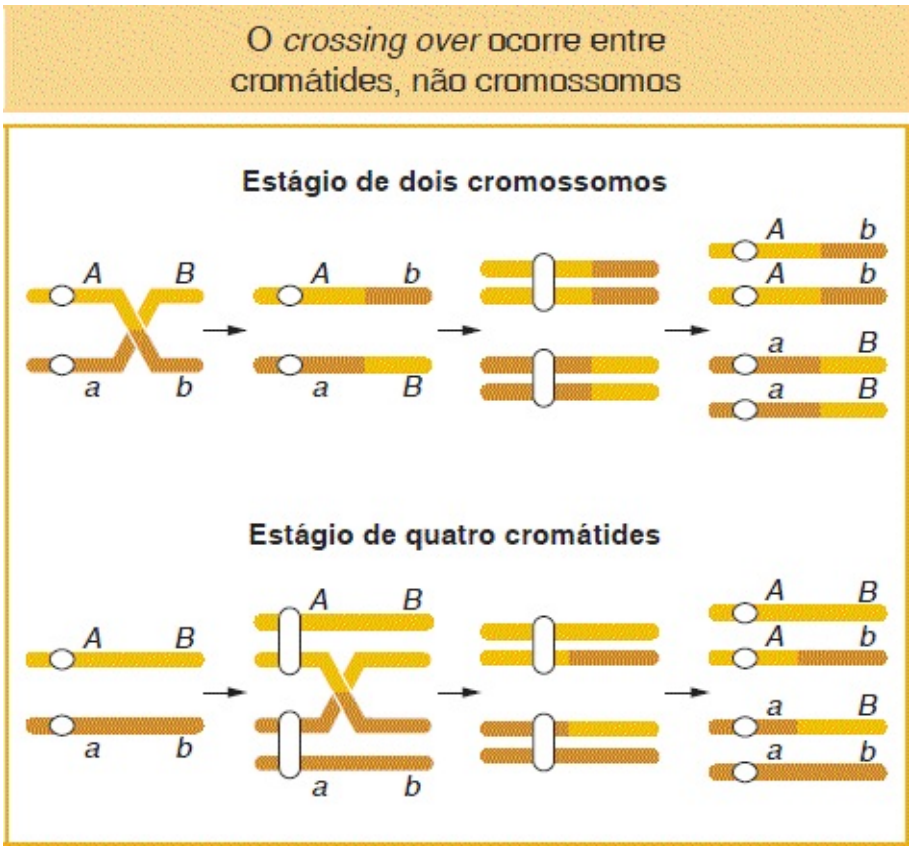


FIGURA 4.4 O *crossing over* ocorre no estágio de quatro cromátides. Tendo em vista que mais de dois produtos diferentes de uma única meiose podem ser observados em algumas tétrades, o *crossing over* não pode ocorrer no estágio de dois filamentos (antes da replicação do DNA). O círculo branco designa a posição do centrômero. Quando as cromátides-irmãs são visíveis, o centrômero parece não replicado.

ABc
AbC
aBC
abc

Essa tétrade é necessariamente explicada por dois *crossovers* nos quais participam *três* cromátides, conforme demonstrado na Figura 4.5 A. Além disso, o tipo de tétrade a seguir demonstra que todas as *quatro* cromátides podem participar no *crossing over* na mesma meiose (Figura 4.5 B):

ABc
Abc

aBC
abC

Portanto, em relação a qualquer par de cromossomos homólogos, duas, três ou quatro cromátides podem participar em eventos de *crossing over* em um único meiócito. Entretanto, observe que qualquer *crossover* único ocorre entre duas cromátides.

Você pode estar se perguntando sobre os *crossovers* entre cromátides-*irmãs*. Eles ocorrem, embora sejam raros, e não produzem novas combinações alélicas, motivo pelo qual normalmente não são considerados.

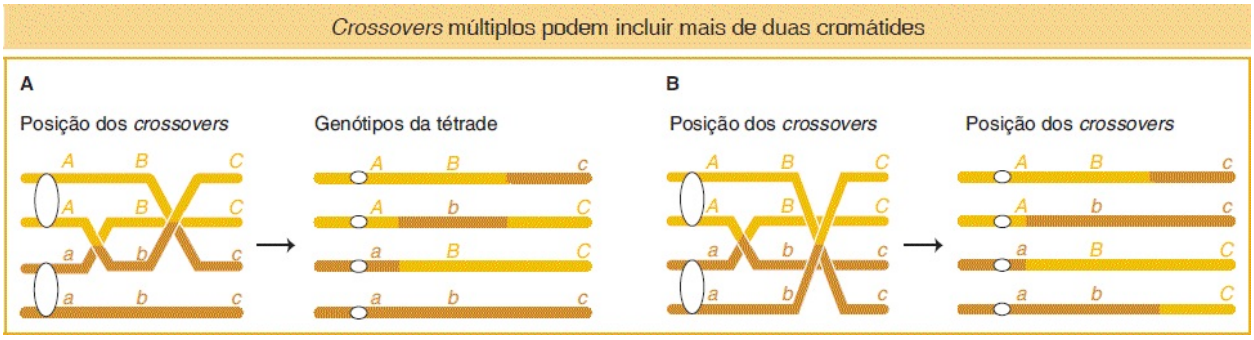


FIGURA 4.5 *Crossovers* duplos podem incluir: (**A**) três cromátides ou (**B**) quatro cromátides.

4.2 Mapeamento por frequência de recombinantes

A frequência de recombinantes produzidos por meio de *crossing over* é a chave para o mapeamento cromossômico. A análise de tétrades em fungos demonstrou que, em relação a quaisquer dois genes ligados específicos, os *crossovers* ocorrem entre eles em alguns dos meiócitos, mas não em todos (Figura 4.6). Quanto mais distantes estiverem os genes, maior será a probabilidade de ocorrência de um *crossover* e mais alta será a proporção de produtos recombinantes. Portanto, a proporção de recombinantes é uma indicação da distância que separa dois *loci* gênicos em um mapa cromossômico.

Conforme declarado anteriormente em relação aos dados de Morgan, a frequência de recombinantes era significativamente inferior a 50%, especificamente 10,7%. A Figura 4.7 demonstra a situação geral em relação à

ligação na qual os recombinantes são menos que 50%. As frequências de recombinantes em relação a diferentes genes ligados variam de 0 a 50%, dependendo da sua proximidade. Quanto mais distantes estiverem os genes, mais suas frequências de recombinantes se aproximam de 50% e, nos referidos casos, não se pode decidir se os genes estão ligados ou se estão em cromossomos diferentes. O que dizer sobre as frequências de recombinantes superiores a 50%? A resposta é que as referidas frequências *nunca* são observadas, conforme será comprovado posteriormente.

Observe na Figura 4.6 que um único *crossover* gera dois produtos recombinantes recíprocos, o que explica o motivo pelo qual as classes de recombinantes recíprocos em geral são aproximadamente iguais em frequência. O corolário desse ponto é que os dois tipos não recombinantes parentais também têm de ser iguais em frequência, conforme também observado por Morgan.

Unidades de mapa

O método básico do mapeamento de genes com a utilização das frequências de recombinantes foi planejado por um aluno de Morgan. Na medida em que Morgan estudava mais e mais genes ligados, ele observou que a proporção da progênie recombinante variava consideravelmente, dependendo de quais genes ligados estavam sendo estudados, e ele acreditava que a referida variação na frequência de recombinantes poderia de algum modo indicar as distâncias reais que separam os genes nos cromossomos. Morgan atribuiu a quantificação desse processo a um aluno de graduação, Alfred Sturtevant, que também se tornou um dos grandes geneticistas. Morgan pediu a Sturtevant que tentasse compreender os dados sobre o *crossing over* entre diferentes genes ligados. Uma noite, Sturtevant desenvolveu um método para mapear genes que ainda é utilizado atualmente. Nas palavras do próprio Sturtevant, “Na parte final de 1911, em uma conversa com Morgan, subitamente percebi que as variações na força de ligação, já atribuídas por Morgan às diferenças na separação espacial dos genes, ofereciam a possibilidade de determinação das sequências na dimensão linear de um cromossomo. Fui para casa e passei a maior parte da noite (tendo negligenciado minha tarefa de casa da graduação) produzindo o primeiro mapa cromossômico”.

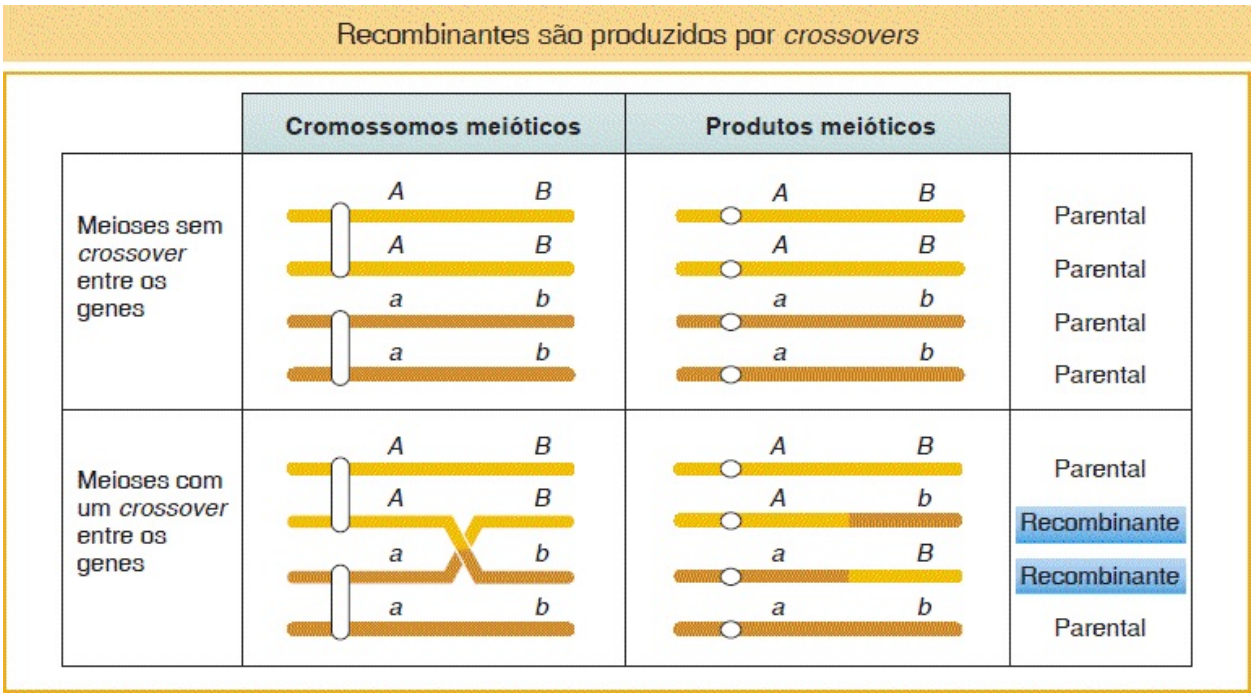


FIGURA 4.6 Os recombinantes surgem a partir de meioses nas quais ocorre um *crossover* entre cromátides não irmãs.

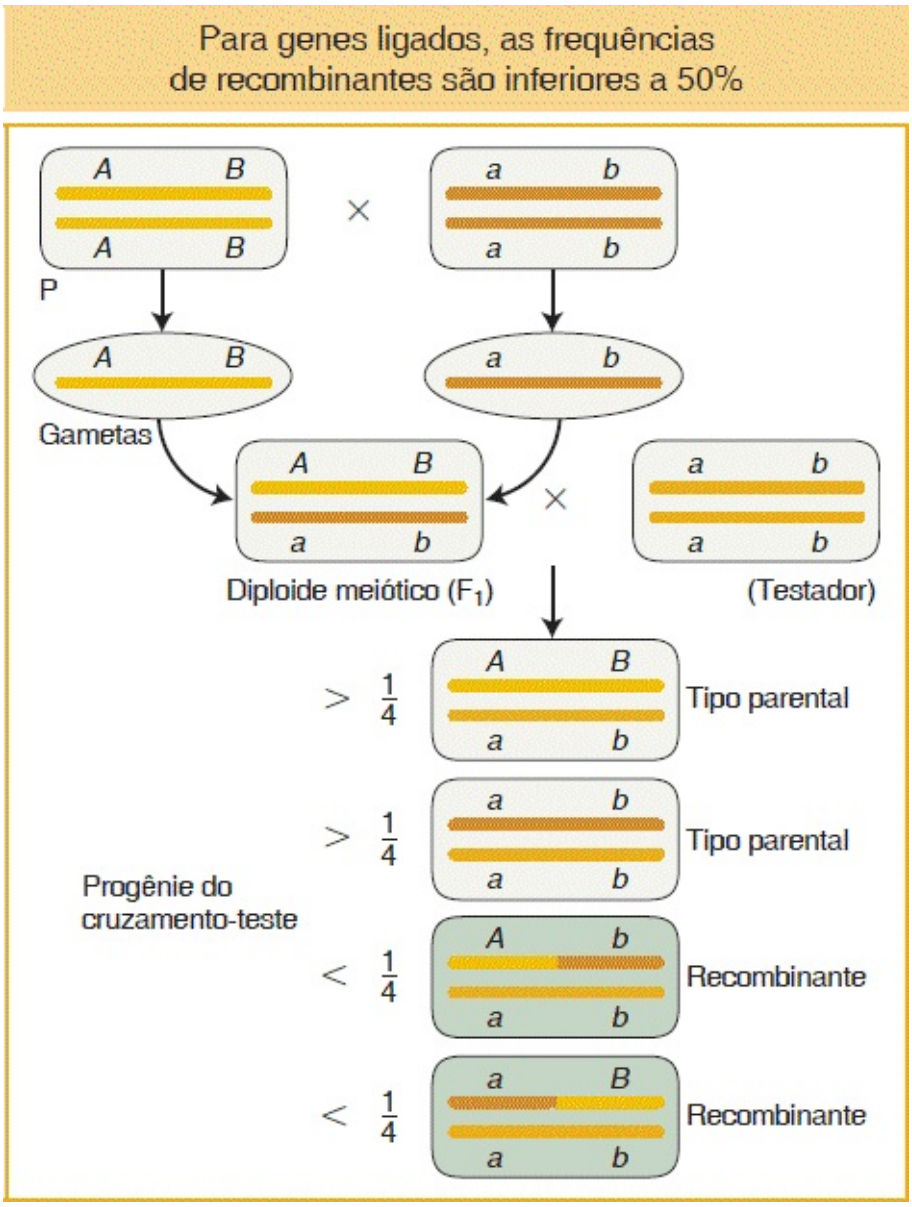


FIGURA 4.7 Um cruzamento-teste revela que as frequências de recombinantes que têm origem dos *crossovers* entre genes ligados são inferiores a 50%.

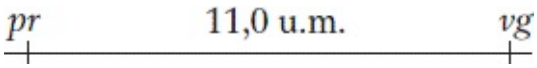
Como um exemplo da lógica de Sturtevant, considere os resultados do cruzamento-teste de Morgan com os genes *pr* e *vg*, a partir dos quais ele calculou uma frequência de recombinantes de 10,7%. Sturtevant sugeriu que podemos utilizar essa porcentagem de recombinantes como um índice quantitativo da distância linear entre dois genes em um mapa genético, ou **mapa de ligação**, conforme por vezes ele é denominado.

A ideia básica aqui é consideravelmente simples. Imagine dois genes específicos posicionados a uma determinada distância fixa. Agora imagine o *crossing over* aleatório ao longo dos homólogos pareados. Em algumas meioses, cromátides não irmãs realizam o *crossover* aleatoriamente na região cromossômica entre esses genes; a partir dessas meioses, são produzidos recombinantes. Em outras divisões meióticas, não existem *crossovers* entre esses genes; nenhum recombinante resulta dessas meioses. (Ver Figura 4.6 para uma ilustração diagramática.) Sturtevant postulou uma proporcionalidade bruta: quanto maior a distância entre os genes ligados, maior a chance de *crossovers* na região entre os genes e, portanto, maior a proporção de recombinantes que será produzida. Assim, ao determinar a frequência de recombinantes, podemos obter uma medida da distância entre os genes no mapa. De fato, Sturtevant definiu uma **unidade de mapa genético (u.m.)** como a distância entre genes em relação aos quais 1 produto da meiose em 100 é recombinante. Por exemplo, a **frequência de recombinantes (FR)** de 10,7% obtida por Morgan é definida como 10,7 u.m. Uma unidade de mapa por vezes é denominada um **centimorgan (cM)**, em homenagem a Thomas Hunt Morgan.

Esse método produz um mapa linear correspondente à linearidade cromossômica? Sturtevant previu que, em um mapa linear, se 5 unidades de mapa (5 u.m.) separam os genes *A* e *B*, e 3 u.m. separam os genes *A* e *C*, então a distância que separa *B* e *C* deve ser ou 8 ou 2 u.m. (Figura 4.8). Sturtevant observou que era o caso da sua previsão. Em outras palavras, sua análise sugeriu fortemente que os genes estão arranjados em alguma ordem linear, tornando as distâncias de mapa aditivas. (Existem algumas exceções, porém não insignificantes, conforme veremos posteriormente.) Tendo em vista que sabemos a partir da análise molecular que um cromossomo é uma molécula de DNA única com os genes arranjados ao longo dela, atualmente para nós não é surpresa saber que os mapas com base na recombinação são lineares, tendo em vista que refletem um arranjo linear de genes.

Como um mapa é representado? Como um exemplo, em *Drosophila*, o *locus* do gene da cor dos olhos e o *locus* do gene do comprimento das asas estão a uma distância de aproximadamente 11 u.m., conforme mencionado anteriormente. A

relação normalmente é diagramada do seguinte modo:



Em geral, fazemos referência ao *locus* desse gene da cor dos olhos de modo abreviado como o "*locus pr*", por ter sido o primeiro alelo mutante descoberto, mas queremos dizer o local no cromossomo onde *qualquer* alelo desse gene será observado, mutante ou do tipo selvagem.

Conforme declarado nos Capítulos 2 e 3, a análise genética pode ser aplicada em sentidos opostos. Esse princípio é aplicável às frequências de recombinantes. Em um sentido, as frequências de recombinantes podem ser utilizadas para criar mapas. No outro, quando nos é fornecido um mapa estabelecido com a distância genética em unidades de mapa, podemos prever as frequências da progênie nas diferentes classes. Por exemplo, a distância genética entre os *loci pr* e *vg* na *Drosophila* é de aproximadamente 11 u.m. Assim, conhecendo esse valor, sabemos que haverá 11% de recombinantes na progênie de um cruzamento-teste de uma fêmea di-híbrida heterozigota em conformação cis ($pr\ vg/pr^+\ vg^+$). Esses recombinantes consistirão em dois recombinantes recíprocos de igual frequência: portanto, 5,5% serão $pr\ vg^+$ e 5,5% serão $pr^+\ vg$. Também sabemos que 100 — 11 = 89% serão não recombinantes em duas classes iguais, 44,5% $pr^+\ vg^+$ e 44,5% *pr vg*. (Observe que a contribuição do testador *pr vg* foi ignorada quando escrevemos esses genótipos.)

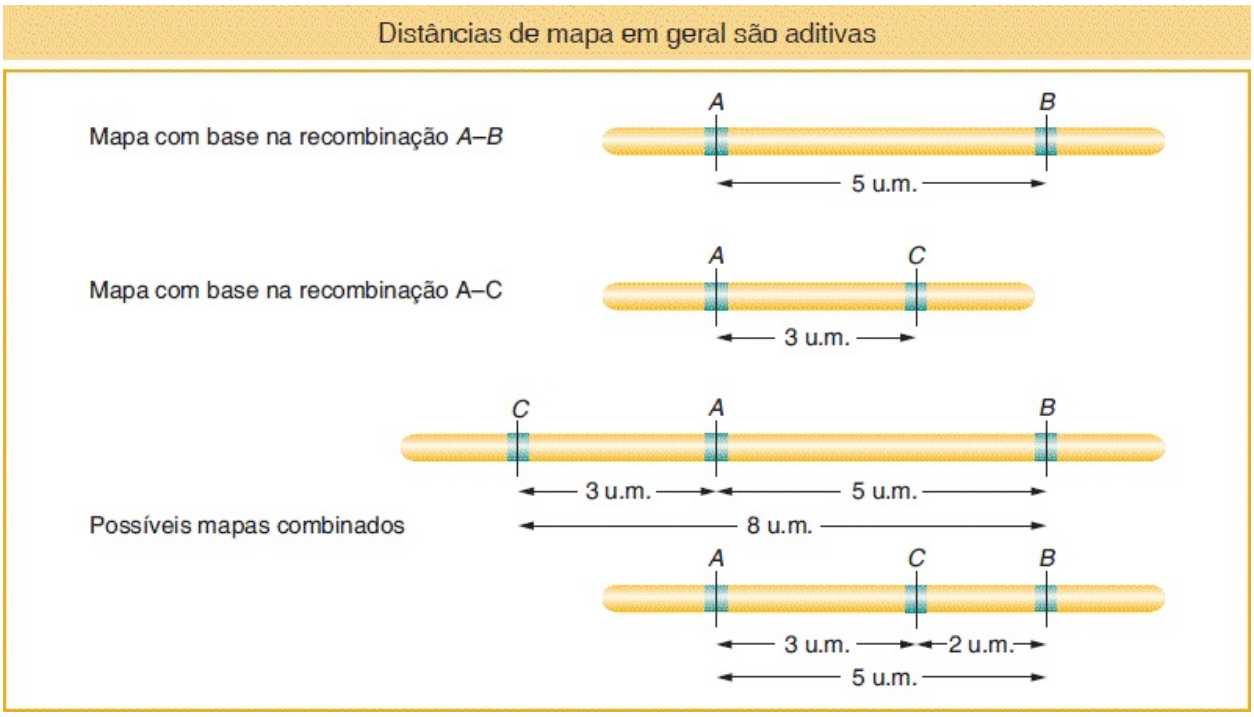


FIGURA 4.8 Uma região cromossômica que contém três genes ligados. Tendo em vista que as distâncias de mapa são aditivas, o cálculo das distâncias A-B e A-C nos deixa com as duas possibilidades demonstradas em relação à distância B-C.

Existe uma forte implicação de que a "distância" em um mapa de ligação seja uma distância física ao longo de um cromossomo, e Morgan e Sturtevant certamente queriam dizer exatamente isso. Mas devemos perceber que o mapa de ligação é uma entidade *hipotética* construída a partir de uma análise puramente genética. O mapa de ligação poderia ter sido derivado até mesmo sem se saber que existiam cromossomos. Além disso, nesse ponto na nossa discussão, não podemos dizer se as "distâncias genéticas" calculadas por meio das frequências de recombinantes representam as distâncias físicas reais nos cromossomos. Entretanto, os mapas físicos demonstraram que as distâncias genéticas são, de fato, aproximadamente proporcionais às distâncias baseadas em recombinação. Existem exceções causadas por *hotspots* ("pontos quentes") de recombinação, locais no genoma em que o *crossing over* ocorre de modo mais frequente do que o habitual. A presença de *hotspots* causa a expansão proporcional de algumas regiões do mapa. Também são conhecidos bloqueios de recombinação, que apresentam o efeito oposto.

Um resumo do modo como os recombinantes do *crossing over* são utilizados no mapeamento está demonstrado na Figura 4.9. Os *crossovers* ocorrem mais ou menos aleatoriamente ao longo do par cromossômico. Em geral, nas regiões mais longas, o número médio de *crossovers* é mais alto e, de acordo com isso, recombinantes são obtidos com mais frequência, o que se traduz em uma distância de mapa mais longa.

> CONCEITO-CHAVE A recombinação entre genes ligados pode ser utilizada para mapear a sua distância em um cromossomo. A unidade de mapeamento (1 u.m.) é definida como uma frequência de recombinantes de 1%.

Cruzamento-teste de três pontos

Até agora, observamos a ligação em cruzamentos de di-híbridos (heterozigotos duplos) com testadores recessivos duplos. O próximo nível de complexidade é um cruzamento de um tri-híbrido (heterozigoto triplo) com um testador recessivo triplo. Esse tipo de cruzamento, denominado **cruzamento-teste de três pontos** ou **cruzamento de três fatores**, é comumente utilizado na análise de ligação. O objetivo é deduzir se três genes estão ligados e, caso afirmativo, deduzir a sua ordem e as distâncias de mapa entre eles.

Observemos um exemplo, também da *Drosophila*. No nosso exemplo, os alelos mutantes são *v* (olhos vermelhos), *cv* (ausência de nervuras nas asas) e *ct* (bordas das asas aparadas, ou cortadas). A análise é feita por meio da realização dos cruzamentos a seguir:

P	$v^+/v^+ \cdot cv/cv \cdot ct/ct \times v/v \cdot cv^+/cv^+ \cdot ct^+/ct^+$
	↓
Gametas	$v^+ \cdot cv \cdot t \quad v \cdot cv^+ \cdot ct^+$
F_1 tri-híbrida	$v/v^+ \cdot cv/cv^+ \cdot ct/ct^+$

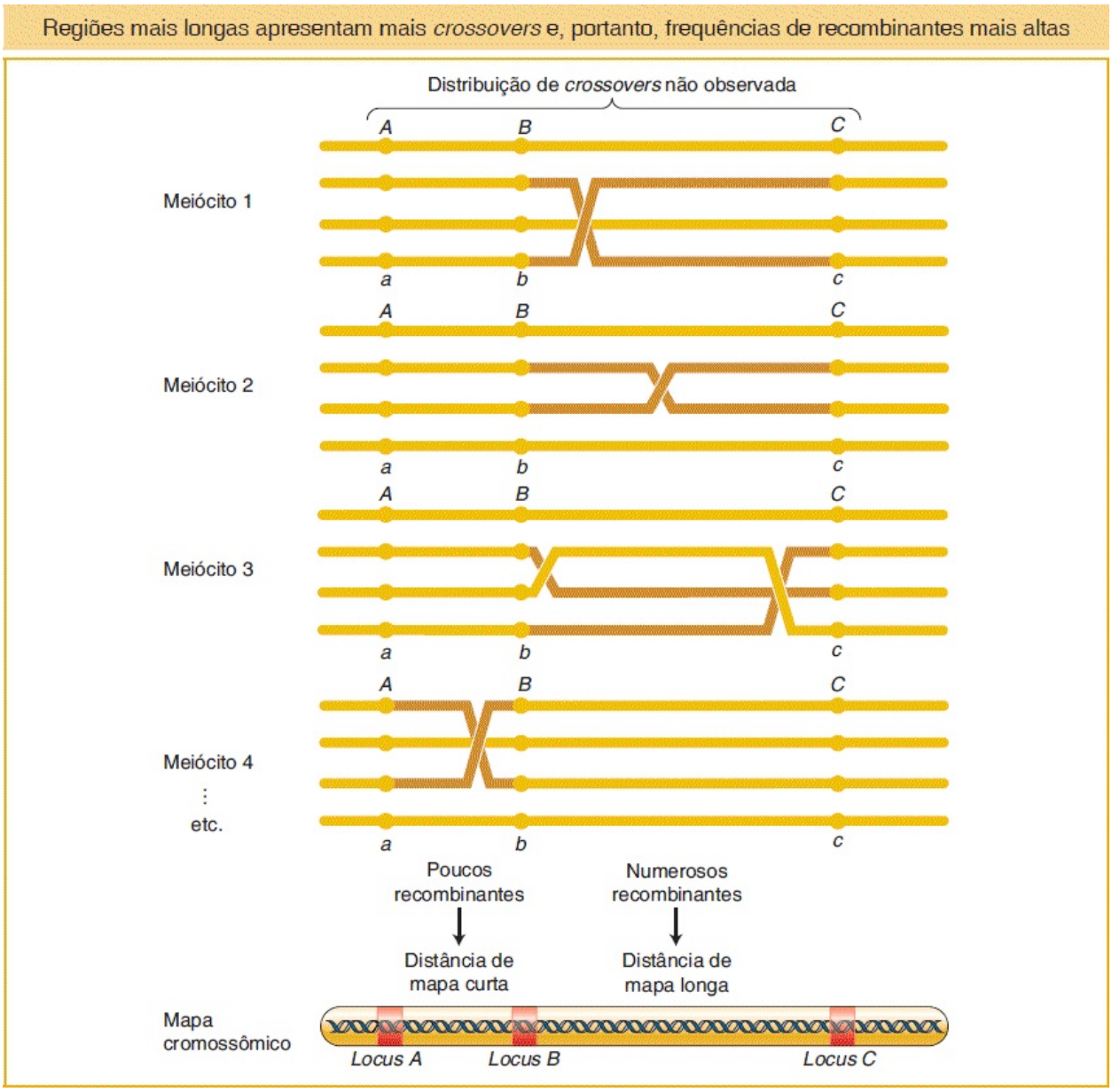


FIGURA 4.9 *Crossovers* produzem cromátides recombinantes cuja frequência pode ser utilizada para mapear os genes em um cromossomo. Regiões mais longas produzem mais *crossovers*. O marrom demonstra os recombinantes em relação àquele intervalo.

Fêmeas tri-híbridas são submetidas ao cruzamento-teste com machos recessivos triplos:

$$\underset{\text{Fêmea tri-híbrida da } F_1}{\female\ v/v^+ \cdot cv/cv^+ \cdot ct/ct^+} \times \underset{\text{Macho testador}}{\male\ v/v \cdot cv/cv \cdot ct/ct}$$

A partir de qualquer tri-híbrido, são possíveis apenas 2 × 2 × 2 = 8 genótipos de gameta. Eles são os genótipos observados na progênie do cruzamento-teste. O quadro a seguir demonstra o número de cada um dos oitos genótipos gaméticos em uma amostra de 1.448 moscas da progênie. As colunas laterais demonstram quais genótipos são recombinantes (R) em relação aos *loci*, considerando-se dois de cada vez. É preciso que sejamos cuidadosos em nossa classificação dos tipos parentais e recombinantes. Observe que os genótipos parentais que entram para os heterozigotos triplos são $v^+ \cdot cv \cdot ct$ e $v \cdot cv^+ \cdot ct^+$; qualquer outra combinação além dessas duas constitui um recombinante.

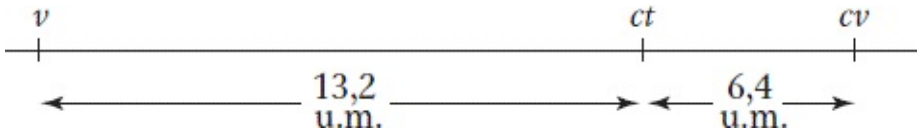
		Recombinantes em relação aos loci		
Gametas		*v* e *cv*	*v* e *ct*	*cv* e *ct*
$v \cdot cv^+ \cdot ct^+$	580			
$v^+ \cdot cv \cdot ct$	592			
$v \cdot cv \cdot ct^+$	45	R		R
$v^+ \cdot cv^+ \cdot ct$	40	R		R
$v \cdot cv \cdot ct$	89	R	R	
$v^+ \cdot cv^+ \cdot ct^+$	94	R	R	
$v \cdot cv^+ \cdot ct$	3		R	R
$v^+ \cdot cv \cdot ct^+$	5		R	R
	1.448	268	191	93

Analisemos os *loci*, dois por vez, iniciando com os *loci v* e *cv*. Em outras palavras, observamos apenas as primeiras duas colunas sob “Gametas” e cobrimos a terceira coluna. Tendo em vista que os parentais em relação a esse par de *loci* são $v^+ \cdot cv$ e $v \cdot cv^+$, sabemos que os recombinantes são, por definição, $v \cdot cv$ e $v^+ \cdot cv^+$. Existem 45 + 40 + 89 + 94 = 268 desses recombinantes. De um total de 1.448 moscas, esse número fornece uma FR de 18,5%.

Em relação aos *loci v* e *ct*, os recombinantes são $v \cdot ct$ e $v^+ \cdot ct^+$. Existem 89 + 94 + 3 + 5 = 191 desses recombinantes entre 1.448 moscas e, assim a FR = 13,2%.

Em relação a *ct* e *cv*, os recombinantes são $cv \cdot ct^+$ e $cv^+ \cdot ct$. Existem 45 + 40 + 3 + 5 = 93 desses recombinantes entre as 1.448 e, assim, a FR = 6,4%.

Claramente, todos os *loci* estão ligados, tendo em vista que os valores das FR são todos consideravelmente inferiores a 50%. Tendo em vista que os *loci v* e *cv* apresentam o maior valor da FR, eles têm de ser os mais distantes. Assim, o *locus ct* encontra-se obrigatoriamente entre eles. Pode ser desenhado um mapa como segue:



O cruzamento-teste pode ser reescrito como segue, agora que sabemos o arranjo de ligação:

v^+ *ct cv*/*v* ct^+ cv^+ × *v ct cv*/*v ct cv*

Aqui, observe diversos pontos importantes. Primeiramente, deduzimos uma ordem dos genes que é diferente daquela utilizada em nossa lista dos genótipos da progênie. Tendo em vista que o ponto do exercício era determinar a relação de ligação desses genes, a lista original era de necessidade arbitrária; a ordem simplesmente não era conhecida antes de os dados serem analisados. Daqui por diante, os genes têm de ser escritos na ordem correta.

Em segundo lugar, estabelecemos definitivamente que *ct* está entre *v* e *cv*. No diagrama, posicionamos *v* arbitrariamente à esquerda e *cv* à direita, mas o mapa

poderia ser igualmente bem-desenhado com o posicionamento invertido desses *loci*.

Em terceiro lugar, observe que os mapas de ligação meramente mapeiam os *loci* em relação uns aos outros, com a utilização das unidades de mapa padrão. Não sabemos onde os *loci* estão em um cromossomo — ou até mesmo em qual cromossomo específico eles estão. Nas análises subsequentes, na medida em que mais *loci* forem mapeados em relação a esses três, o mapa cromossômico completo seria "concluído".

> CONCEITO-CHAVE Cruzamentos-teste de três pontos (e superiores) possibilitam que os geneticistas avaliem a ligação entre três (ou mais) genes e determinem a ordem dos genes, tudo em um cruzamento.

Por fim, é preciso que as duas distâncias de mapa menores, 13,2 u.m. e 6,4 u.m., somam 19,6 u.m., que é superior a 18,5 u.m., a distância calculada em relação a *v* e *cv*. Por quê? A resposta dessa questão está no modo como tratamos as duas classes de progênie mais raras (que totalizam 8) em relação à recombinação de *v* e *cv*. Agora que temos o mapa, podemos observar que essas duas classes raras são de fato recombinantes duplos, que surgem a partir de dois *crossovers* (Figura 4.10). Entretanto, quando calculamos o valor da FR em relação a *v* e *cv*, não contamos os genótipos *v ct* cv^+ e v^+ ct^+ *cv*; afinal, em relação a *v* e *cv*, eles são combinações parentais (*v* cv^+ e v^+ *cv*). À luz do nosso mapa, entretanto, observamos que esse descuido nos levou a subestimar a distância entre os *loci v* e *cv*. Não apenas deveríamos ter contato as duas classes mais raras, como deveríamos ter contato cada uma delas *duas vezes*, tendo em vista que cada uma representa recombinantes duplos. Portanto, podemos corrigir o valor ao adicionar os números 45 + 40 + 89 + 94 + 3 + 3 + 5 + 5 = 284. Do total de 1.448, essa quantidade é exatamente 19,6%, que é idêntica à soma dos dois valores componentes. (Na prática, não necessitamos desse cálculo, tendo em vista que a soma das duas distâncias mais curtas nos fornece a melhor estimativa da distância geral.)

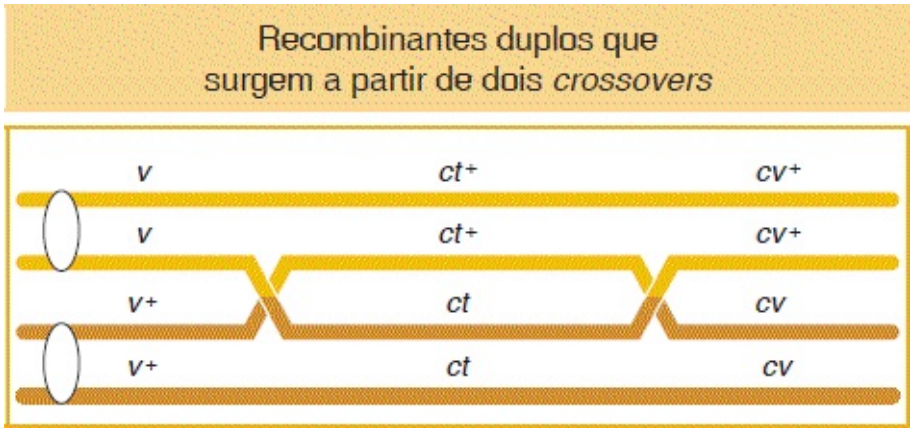


FIGURA 4.10 Exemplo de um *crossover* duplo entre duas cromátides. Observe que um *crossover* duplo produz cromátides recombinantes duplas, que apresentam as combinações de alelos parentais em outros *loci*. A posição do centrômero não pode ser determinada a partir dos dados. Ele foi adicionado para fins de integridade.

Dedução da ordem dos genes por inspeção

Agora que obtivemos alguma experiência com o cruzamento-teste de três pontos, podemos observar novamente a listagem da progênie e verificar que, em relação a tri-híbridos de genes ligados, a *ordem dos genes* normalmente pode ser deduzida por inspeção, sem uma análise da frequência de recombinantes. Tipicamente, em relação aos genes ligados, temos os oito genótipos nas frequências a seguir:

dois com frequência alta

dois com frequência intermediária

dois com frequência intermediária diferente

dois raros

Apenas três ordens de genes são possíveis, cada uma com um gene diferente na posição intermediária. Em geral, é verdadeiro que as classes de recombinantes duplos são as menores, conforme listado aqui por último. Apenas uma ordem é compatível com a formação das classes menores por meio de *crossovers* duplos, conforme demonstrado na Figura 4.11; ou seja, apenas uma ordem fornece recombinantes duplos dos genótipos *v ct* cv^+ e v^+ ct^+ *cv*. Uma regra geral simples para deduzir o gene do meio é que ele é o par de alelos que “mudou” de posição

nas classes de recombinantes duplos.

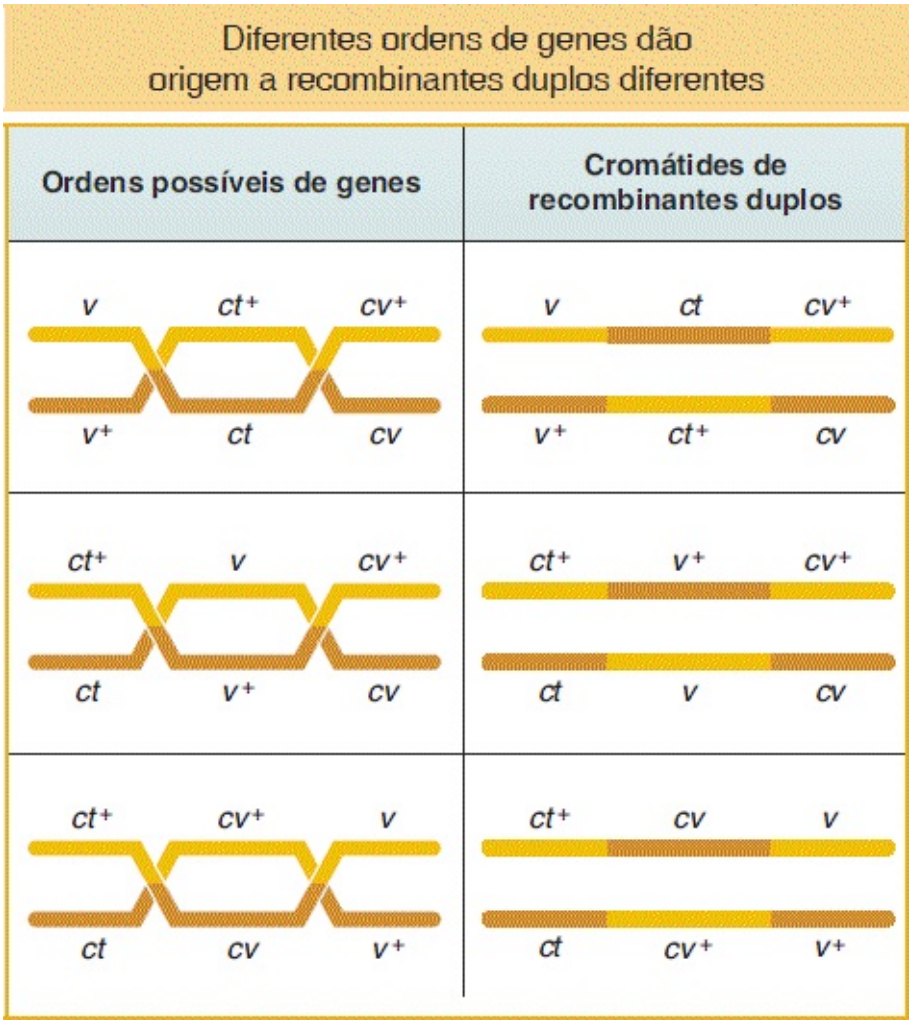


FIGURA 4.11 As três ordens de genes possíveis demonstradas à esquerda produzem os seis produtos de um *crossover* duplo demonstrado à direita. Apenas a primeira possibilidade é compatível com os dados no texto. Observe que apenas as cromátides não irmãs que participam no *crossover* duplo estão demonstradas.

Interferência

Conhecer a existência dos *crossovers* duplos nos permite indagar questões a respeito da sua possível interdependência. Podemos indagar: os *crossovers* em regiões cromossômicas adjacentes são eventos independentes ou um *crossover* em uma região afeta a probabilidade de haver um *crossover* em uma região adjacente? A resposta é que, em geral, os *crossovers* inibem uns aos outros de um modo semelhante a uma interação, denominado **interferência**. As classes de

recombinantes duplos podem ser utilizadas para deduzir a extensão dessa interferência.

A interferência pode ser medida do modo a seguir. Se os *crossovers* nas duas regiões forem independentes, podemos utilizar a regra do produto (ver Capítulo 3) para prever a frequência de recombinantes duplos: aquela frequência será igual ao produto das frequências de recombinantes nas regiões adjacentes. Nos dados da recombinação de *v-ct-cv*, o valor da FR de *v-ct* é 0,132 e o valor de *ct-cv* é 0,064. Assim, se não houver interferência, os recombinantes duplos podem ser esperados na frequência de 0,132 × 0,064 = 0,0084 (0,84%). Na amostra de 1.448 moscas, 0,0084 × 1.448 = 12 recombinantes duplos são esperados. Mas os dados demonstram que apenas 8 foram realmente observados. Se essa deficiência de recombinantes duplos fosse consistentemente observada, ela nos mostraria que as duas regiões não são independentes e sugeriria que a distribuição dos *crossovers* favorece a ocorrência de recombinantes únicos à custa dos duplos. Em outras palavras, existe algum tipo de interferência: um *crossover* reduz a probabilidade de um *crossover* em uma região adjacente.

A interferência é quantificada primeiro por meio do cálculo de um termo denominado **coeficiente de coincidência (c.d.c.)**, que é a proporção de recombinantes duplos observados com relação à esperada. A interferência (I) é definida como 1 — c.d.c. Portanto,

$$I = 1 - \frac{\text{Frequência ou número observado de recombinantes duplos}}{\text{Frequência ou número esperado de recombinantes duplos}}$$

Em nosso exemplo,

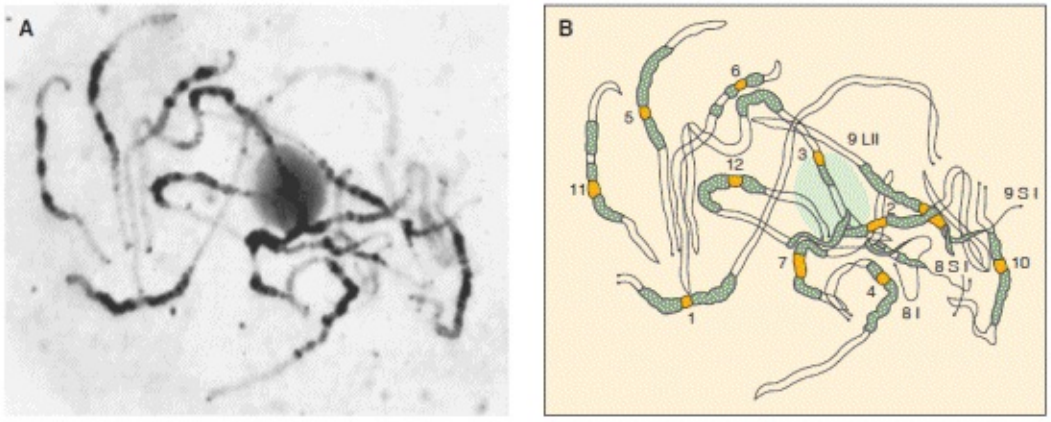
$$I = 1 - \frac{8}{12} = \frac{4}{12} = \frac{1}{3}, \text{ ou } 33\%$$

Em algumas regiões, nunca foram observados quaisquer recombinantes duplos. Nesses casos, c.d.c. = 0 e, assim, I = 1 e a interferência é completa. Valores de interferência em qualquer localização entre 0 e 1 são observados em diferentes regiões e em diferentes organismos.

Você pode ter se perguntado o motivo de sempre utilizarmos fêmeas heterozigotas para cruzamentos-teste em *Drosophila*. A explicação está em uma característica incomum dos machos de *Drosophila*. Quando, por exemplo, machos *pr vg/pr*$^{+}$ *vg*$^{+}$ são cruzados com fêmeas *pr vg/pr vg*, apenas decendentes *pr vg/pr*$^{+}$ *vg*$^{+}$ e *pr vg/pr vg* são recuperadas. Esse resultado demonstra que não existe *crossing over* em machos de *Drosophila*. Entretanto, essa ausência de *crossing over* em um sexo é limitada a determinadas espécies, não é o caso dos machos de todas as espécies (ou em relação ao sexo heterogamético). Em outros organismos, ocorre *crossing over* em machos XY e em fêmeas WZ. O motivo da ausência de *crossing over* em machos de *Drosophila* é que eles apresentam uma prófase I incomum, sem complexos sinaptonêmicos. Incidentalmente, existe uma diferença de recombinação também entre os sexos humanos. Mulheres demonstram frequências de recombinantes mais altas em relação aos mesmos *loci* autossômicos do que os homens.

Com a utilização de uma reiteração das técnicas precedentes baseadas na recombinação, foram produzidos mapas de milhares de genes em relação aos quais foram identificados fenótipos variantes (mutantes). Um exemplo ilustrativo simples do tomate está demonstrado na Figura 4.12. Os cromossomos do tomate estão demonstrados na Figura 4.12 A, a sua numeração na Figura 4.12 B e os mapas genéticos com base na recombinação, na Figura 4.12 C. Os cromossomos estão demonstrados como aparecem sob o microscópio, juntamente com os mapas cromossômicos baseados na análise de ligação de diversos pares alélicos mostrados com seus fenótipos.

Mapa dos 12 cromossomos de tomate



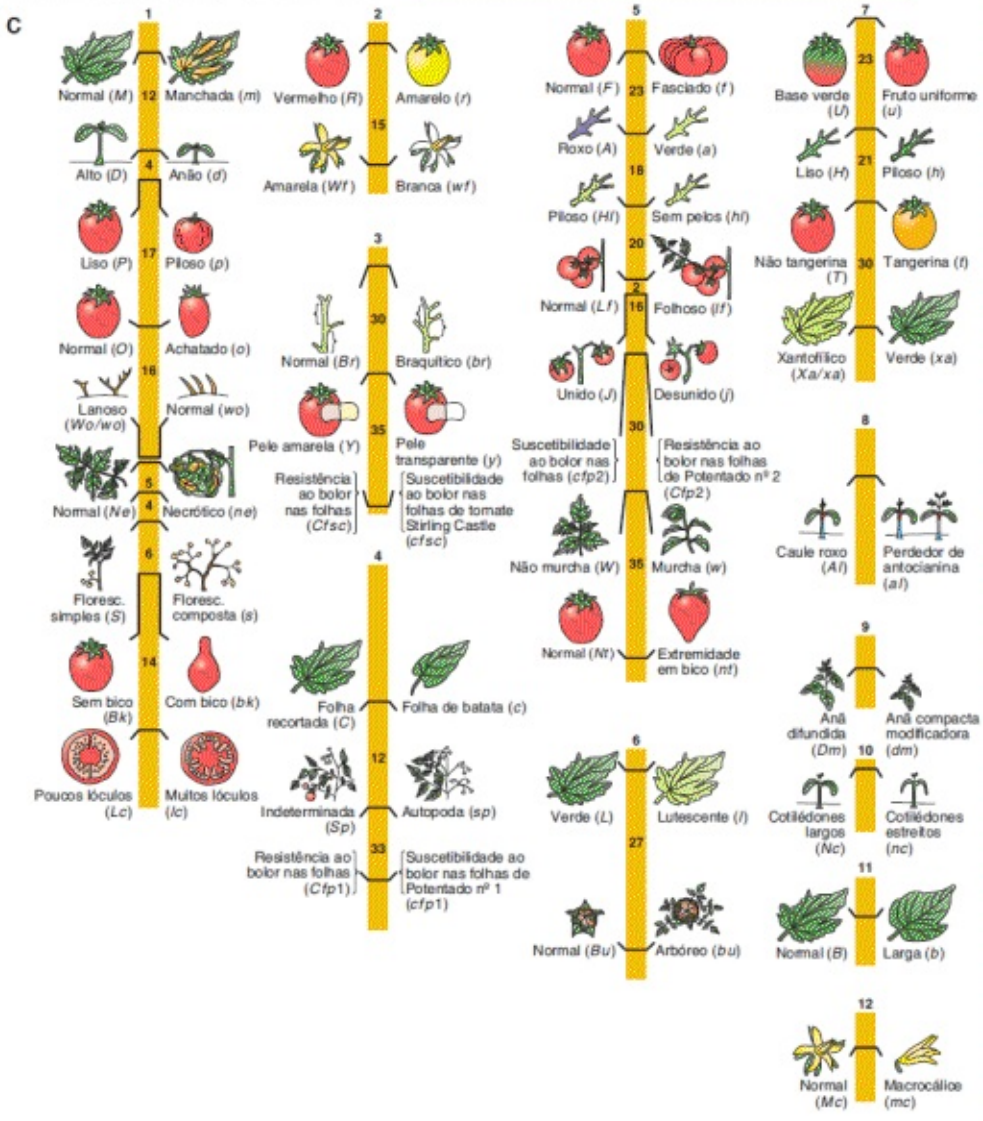


FIGURA 4.12 A. Fotomicrografia de uma prófase I meiótica (paquíteno) das anteras, demonstrando os 12 pares de cromossomos. **B.** Ilustração dos 12 cromossomos demonstrados na parte **A.** Os cromossomos estão identificados por meio do sistema de numeração cromossômica atualmente utilizado. Os centrômeros estão demonstrados em laranja e as regiões flanqueadoras, densamente coradas (heterocromatina), em verde. **C.** Mapa de ligação de 1952. Cada *locus* é ladeado por desenhos dos fenótipos normais e variantes. As distâncias de mapa *interlocus* estão demonstradas em unidades de mapa. (*A* e *B.* de C. M. Rick, "The Tomato", *Scientific American*, 1978; *C.* Dados de L. A. Butler.)

Utilização de proporções como diagnóstico

A análise das proporções é um dos pilares da genética. Até agora no texto, encontramos muitas proporções diferentes, cujas derivações estão difundidas ao longo de diversos capítulos. Tendo em vista que o reconhecimento das proporções e a sua utilização no diagnóstico do sistema genético em estudo fazem parte da genética diária, revisaremos as principais proporções que abrangemos até agora. Elas estão demonstradas na Figura 4.13. Você pode ler as proporções pelas larguras relativas dos boxes coloridos em uma fileira. A Figura 4.13 se refere às autofecundações e aos cruzamentos-teste de mono-híbridos, di-híbridos (com distribuição independente e ligação) e tri-híbridos (também com distribuição independente e ligação de todos os genes). Uma situação não representada é um tri-híbrido no qual apenas dois dos três genes estão ligados; como um exercício, você pode querer deduzir o padrão geral que deveria ser incluído em um referido diagrama a partir dessa situação. Observe que, em relação à ligação, os tamanhos das classes dependem das distâncias de mapa. Um geneticista deduz os estados genéticos desconhecidos em um modo semelhante ao seguinte: "uma proporção de 9:3:3:1 me informa que essa proporção muito provavelmente foi produzida por um di-híbrido autofecundado, no qual os genes estão em cromossomos diferentes".

4.3 Mapeamento com marcadores moleculares

Até agora neste capítulo, mapeamos os *loci* de genes com a utilização dos valores da FR ao contar os fenótipos visíveis produzidos pelos diversos alelos envolvidos. Entretanto, também existem diferenças no DNA entre dois

cromossomos que não produzem fenótipos visivelmente diferentes, seja em virtude dessas diferenças no DNA não estarem localizadas em genes ou por estarem localizadas em genes, mas não alterarem o produto proteico. As referidas diferenças na sequência podem ser consideradas alelos moleculares ou **marcadores moleculares**. Seus *loci* podem ser mapeados por meio dos valores de FR do mesmo modo que os alelos que produzem fenótipos visíveis. Os marcadores moleculares são extremamente numerosos e, portanto, são muito úteis como marcos genômicos, que podem ser utilizados para localizar os genes de interesse.

Os dois tipos principais de marcadores moleculares utilizados no mapeamento são os polimorfismos de nucleotídio único e os polimorfismos de comprimento de sequência simples.

Polimorfismos de nucleotídio único

O sequenciamento demonstrou que, conforme esperado, as sequências genômicas dos indivíduos em uma espécie são, em sua maior parte, idênticas. Por exemplo, comparações das sequências de diferentes indivíduos revelaram que somos idênticos em aproximadamente 99,9%. Quase toda a diferença de 0,1% ocorre com base em diferenças de um único nucleotídio. Como um exemplo, em um indivíduo, uma sequência localizada pode ser:

….AA**G**GCTCAT….
….TT**C**CGAGTA….

e, em outro, pode ser:

….AA**A**GCTCAT….
….TT**T**CGAGTA….

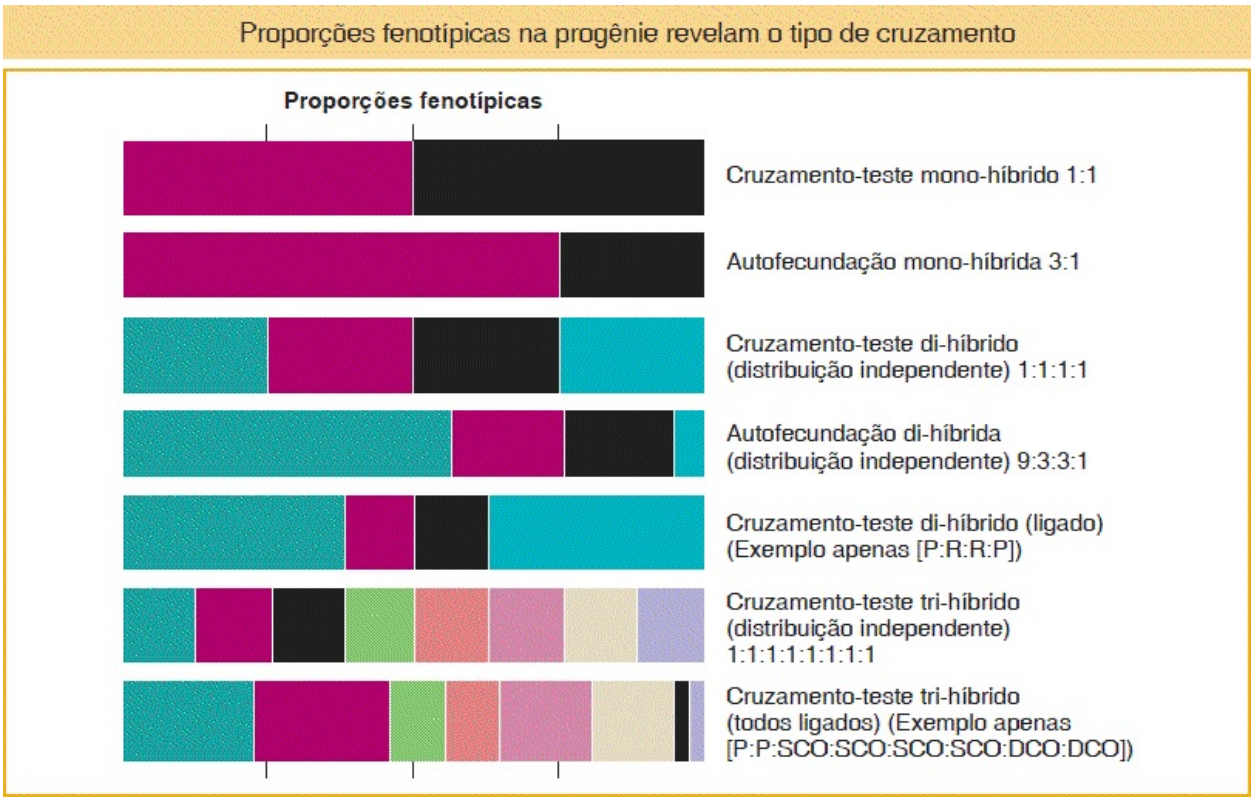


FIGURA 4.13 P = Parental; R = Recombinante; SCO = *Crossover* único; DCO = *Crossover* duplo.

Além disso, observa-se que uma grande proporção dessas sequências localizadas é polimórfica, o que significa que ambos os "alelos" moleculares são consideravelmente comuns na população. Em geral, as referidas diferenças entre os indivíduos são denominadas **polimorfismos de nucleotídio único**, abreviadas como **SNP** (do inglês, *single nucleotide polymorphisms*). Em seres humanos, acredita-se que existam aproximadamente 3 milhões de SNP distribuídos mais ou menos aleatoriamente, a uma frequência de 1 em cada 300 a 1.000 bases.

Alguns desses SNP estão localizados nos genes; muitos deles, não. No Capítulo 2, vimos casos nos quais a alteração em um único par de nucleotídios pode produzir um novo alelo, causando um fenótipo mutante. Os dois pares de nucleotídios, do tipo selvagem e mutante, são exemplos de um SNP. Entretanto, a maior parte dos SNP não produz fenótipos diferentes, seja em virtude de não estarem localizados em um gene ou em virtude de estarem localizados em um gene, mas ambas as versões do gene produzem o mesmo produto proteico.

Existem dois modos de detectar um SNP. O primeiro é sequenciar um segmento

de DNA em cromossomos homólogos e comparar os segmentos homólogos para localizar as diferenças. Um segundo modo é possível no caso de SNP localizados em um sítio-alvo de uma enzima de restrição: esses SNP são **polimorfismos de comprimento de fragmento de restrição** (**RFLP**; do inglês, *restriction fragment length polymorphisms*). Nos referidos casos, haverá dois "alelos" de RFLP, ou morfos, um dos quais apresenta o alvo da enzima de restrição, enquanto o outro não apresenta. A enzima de restrição cortará o DNA no SNP que contém o alvo e ignorará o outro SNP. Os SNP são, assim, detectados como bandas diferentes em um gel de eletroforese. Os sítios de RFLP podem estar entre ou dentro dos genes.

Polimorfismos de comprimento de sequência simples

Uma das surpresas da análise genômica molecular é que a maior parte dos genomas contém uma grande quantidade de DNA repetitivo. Além disso, existem muitos tipos de DNA repetitivo. Em uma extremidade do espectro estão as múltiplas repetições adjacentes de sequências de DNA curtas e simples. A origem dessas repetições não está clara, mas a característica que as torna úteis é que, em indivíduos diferentes, com frequência existem números diferentes de cópias. Portanto, essas repetições são denominadas **polimorfismos de comprimento de sequência simples** (**SSLP**; do inglês, *simple sequence length polymorphisms*). Por vezes elas também são denominadas **repetições em tandem de número variável**, ou **VNTR** (do inglês, *variable number tandem repeats*).

Os SSLP comumente apresentam alelos múltiplos; até 15 alelos foram encontrados em relação a um *locus* de SSLP. Por consequência, por vezes quatro alelos (dois de cada genitor) podem ser seguidos em um heredograma. Dois tipos de SSLP são úteis no mapeamento e em outras análises do genoma: os marcadores minissatélites e microssatélites. (A palavra *satélite* nessa conexão se refere à observação de que, quando o DNA genômico é isolado e fracionado com a utilização de técnicas físicas, as sequências repetitivas com frequência formam uma fração que está fisicamente separada do restante; ou seja, é uma fração satélite no sentido em que está separada da maior parte.)

Marcadores minissatélites.

Um **marcador minissatélite** tem por base a variação

no número de repetições em tandem de uma unidade de repetição com comprimento de 15 a 100 nucleotídios. Em seres humanos, o comprimento total da unidade é de 1 a 5 kb. Os *loci* de minissatélites que apresentam a mesma unidade de repetição, mas números diferentes de repetições, estão dispersos por todo o genoma.

Marcadores microssatélite. Um **marcador microssatélite** tem por base números variáveis de repetições em tandem de uma sequência ainda mais simples, em geral um pequeno número de nucleotídios, tal como um dinucleotídio. O tipo mais comum é uma repetição de CA e do seu complemento GT, como no exemplo a seguir:

5' C-A-C-A-C-A-C-A-C-A-C-A-C-A-C-A 3'
3' G-T-G-T-G-T-G-T-G-T-G-T-G-T-G-T 5'

Detecção de polimorfismos de comprimento de sequência simples

Os polimorfismos de comprimento de sequência simples são detectados aproveitando-se da vantagem de que regiões homólogas que contêm números diferentes de repetições em tandem serão de diferentes comprimentos. Um procedimento comumente utilizado para revelar essas diferenças é utilizar as regiões flanqueadoras como *primers* em uma análise de PCR (ver Capítulo 10). A PCR replica as sequências de DNA até que elas estejam disponíveis em quantidade suficiente para análises adicionais. Os diferentes comprimentos dos produtos amplificados pela PCR podem ser detectados por meio das diferentes mobilidades das sequências em um gel de eletroforese. No caso dos minissatélites, os padrões produzidos no gel por vezes são denominados **impressões digitais do DNA**. (Tais impressões digitais são altamente individualizadas e, portanto, apresentam grande valor para a ciência forense, conforme detalhado no Capítulo 18.)

Análise de recombinação com a utilização de marcadores moleculares

Quando mapeamos a posição de um gene cujos fenótipos são determinados pela diferença de um único nucleotídio, estamos efetivamente mapeando um SNP. A mesma técnica utilizada para mapear os *loci* gênicos também pode ser utilizada para mapear os SNP que não determinam um fenótipo.

Suponha que um indivíduo apresente um par de bases GC, digamos, na posição 5.658 no DNA de um cromossomo e um AT na posição 5.658 no outro cromossomo. Um referido indivíduo é um heterozigoto molecular ("AT/GC") em relação àquela posição do DNA. Esse fato é útil para o mapeamento, tendo em vista que um heterozigoto molecular ("AT/GC") pode ser mapeado justamente do mesmo modo que um heterozigoto fenotípico *A/a*. O *locus* de um heterozigoto molecular pode ser inserido em um mapa cromossômico por meio da análise da frequência de recombinação exatamente do mesmo modo como é inserido o *locus* de alelos heterozigotos "fenotípicos". Esse princípio é mantido ainda que a variação normalmente seja uma diferença silenciosa (talvez não em um gene).

Atuando como importantes "marcos" no mapa, os marcadores moleculares são úteis para orientar o pesquisador em uma investigação para encontrar um gene de interesse. Para compreender esse ponto, considere os marcos reais: eles são de pouco interesse por si próprios, mas são muito úteis para lhe informar quão próximo você se encontra do seu destino. Em um exemplo genético específico, presumiremos que desejamos conhecer a posição no mapa de um gene de doença em camundongos, talvez como um modo de detectá-lo em sua sequência de DNA. Realizamos uma diversidade de cruzamentos. Em cada um dos casos, cruzamos um indivíduo portador do gene da doença com um indivíduo que carreia um de uma gama de diferentes marcadores moleculares, cujas posições no mapa já são conhecidas. Com a utilização da PCR, os genitores e a progênie são classificados em relação aos marcadores moleculares de posição no mapa conhecida e, em seguida, é realizada a análise de recombinação para verificar se o gene de interesse está ligado a qualquer um deles. O resultado desses cruzamentos pode revelar que o gene da doença está à distância de 2 u.m. de um desses marcadores, que denominaremos M. Portanto, o procedimento nos forneceu uma localização aproximada do gene da doença no cromossomo. A localização do gene da doença humana fibrose cística foi originalmente descoberta por meio da sua ligação com

marcadores moleculares que se sabia estarem localizados no cromossomo 7. Essa descoberta levou ao isolamento e ao sequenciamento do gene, resultando na descoberta adicional de que ele codifica a proteína atualmente denominada *reguladora de condutância transmembrana da fibrose cística* (CFTR; do inglês, *cystic fibrosis transmembrane condutance regulator*). O gene da doença de Huntington também foi localizado desse modo, levando à descoberta de que ele codifica uma proteína muscular denominada *huntingtina*.

O procedimento experimental em relação a um exemplo hipotético pode ser como segue. Considere que *A* e *a* sejam os alelos do gene da doença e M1 e M2 sejam os alelos de um *locus* de marcador molecular específico. Presuma que o cruzamento é *A/a* · M1/M2 × *a/a* · M1/M1, um tipo de cruzamento-teste. A progênie seria classificada primeiramente em relação aos fenótipos *A* e *a*, e em seguida o DNA seria extraído de cada indivíduo e sequenciado, ou de outro modo avaliado, a fim de determinar os alelos moleculares. Presuma que obtemos os resultados a seguir:

A/a · M1/M1 49%	*A/a* · M2/M1 1%
a/a · M2/M1 49%	*a/a* · M1/M1 1%

Esses resultados nos informam que o cruzamento-teste tem obrigatoriamente a conformação a seguir:

A M1/*a* M2 × *a* M1/*a* M1

e os dois genótipos da progênie à direita na lista têm de ser recombinantes, fornecendo uma distância de mapa de 2 u.m. entre o *locus A/a* e o *locus* molecular M1/M2. Portanto, agora sabemos a localização geral do gene no genoma e podemos estreitar ainda mais a sua localização com abordagens mais meticulosas. Além disso, diferentes marcadores moleculares podem ser mapeados entre si, criando-se um mapa que pode atuar como uma série de trampolins no caminho até algum gene com um fenótipo de interesse.

Embora o mapeamento de marcadores moleculares com a utilização do que efetivamente são cruzamentos-teste seja o tipo mais simples de análise

informativa, em muitas análises (tais como aquelas em seres humanos) os marcadores moleculares não podem ser mapeados com a utilização de um cruzamento-teste. Entretanto, tendo em vista que cada alelo molecular apresenta a sua própria assinatura, os produtos recombinantes e não recombinantes podem ser identificados a partir de qualquer meiose, até mesmo em cruzamentos que não são cruzamentos-teste. Uma referida análise está diagramada na Figura 4.14.

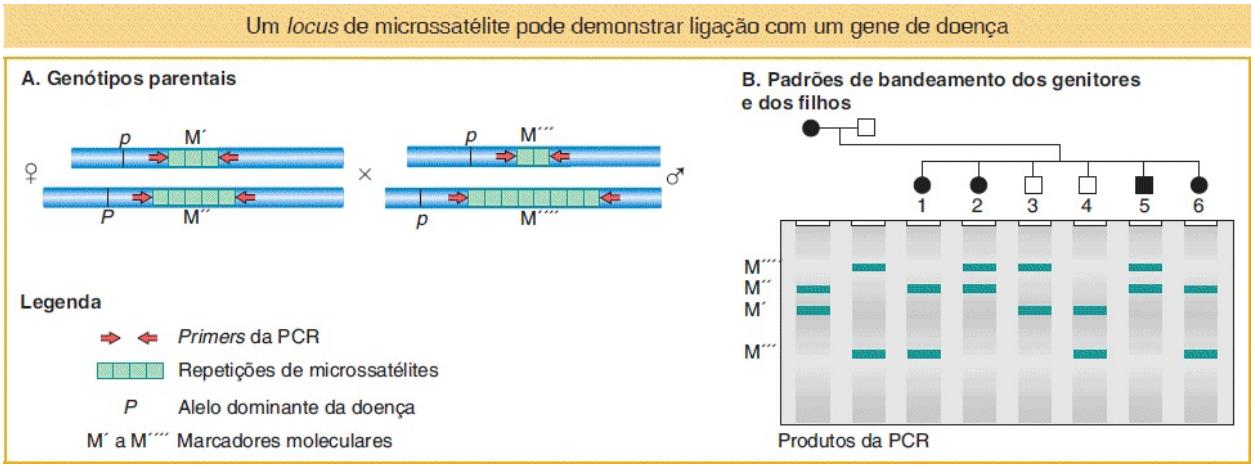


FIGURA 4.14 Um padrão de bandeamento de PCR é demonstrado em relação a uma família com seis filhos. Tal padrão está interpretado na parte superior da ilustração com a utilização de quatro "alelos" microssatélites de tamanhos diferentes, M' a M''. Um desses marcadores (M") provavelmente está ligado em configuração cis ao alelo P da doença. (Nota: este não é um cruzamento-teste, ainda que seja informativo a respeito da ligação.)

A Figura 4.15 contém alguns dados reais que demonstram como os marcadores moleculares podem contribuir para o mapa de um cromossomo humano. Você pode observar que o número de marcadores moleculares mapeados excede em muito o número de genes mapeados com fenótipos mutantes. Observe que os SNP, em virtude de sua densidade ainda mais alta, não podem ser representados em um mapa de cromossomo inteiro, tal como aquele na Figura 4.15, uma vez que haveria milhares deles. Um centimorgan (1 u.m.) do DNA humano é um segmento enorme, estimado como 1 megabase (1 Mb = 1 milhão de pares de bases, ou 1.000 kb). Portanto, você pode verificar a necessidade de marcadores moleculares próximos para uma análise meticulosa de distâncias menores. Observe que o DNA equivalente a 1 u.m. varia muito entre as espécies; por

exemplo, no parasita da malária, *Plasmodium falciparum*, 1 u.m. = 17 kb.

> CONCEITO-CHAVE Os *loci* de qualquer heterozigosidade do DNA podem ser mapeados e utilizados como marcadores cromossômicos moleculares ou marcos.

4.4 Mapeamento de centrômeros com tétrades lineares

Centrômeros não são genes, mas regiões do DNA das quais a reprodução ordenada dos organismos vivos depende totalmente e, portanto, são de grande interesse em genética. Na maior parte dos eucariotos, a análise de recombinação não pode ser utilizada para mapear os *loci* dos centrômeros, tendo em vista que eles não demonstram heterozigosidade que lhes possibilite serem utilizados como marcadores. Entretanto, nos fungos que produzem tétrades lineares (ver Capítulo 3), os centrômeros *podem* ser mapeados. Utilizaremos o fungo *Neurospora* como um exemplo. Relembre que, em fungos haploides, tais como *Neurospora*, os núcleos haploides de cada genitor se fundem para formar um diploide temporário, que é submetido a divisões meióticas ao longo do eixo longitudinal do asco e, assim, cada meiócito produz um arranjo linear de oito ascósporos, denominado **óctade**. Esses oito ascósporos constituem os quatro produtos da meiose (uma tétrade), mais uma mitose pós-meiótica.

Marcadores fenotípicos e moleculares mapeados no cromossomo humano 1

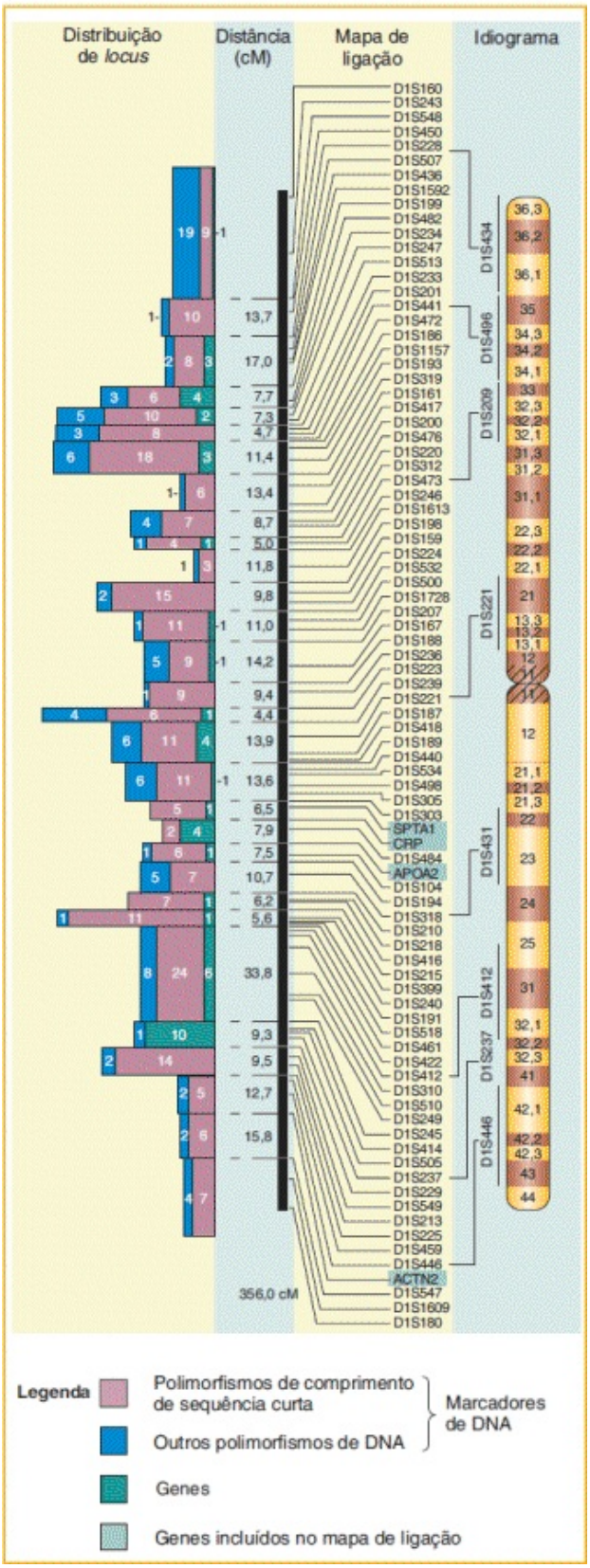


FIGURA 4.15 O diagrama demonstra a distribuição de todas as diferenças genéticas que haviam sido mapeadas no cromossomo 1 na ocasião em que este diagrama foi desenhado. Alguns marcadores são genes de fenótipos conhecidos (seus números estão sombreados em verde), mas a maior parte é de marcadores de DNA polimórficos (os números sombreados em malva e azul representam duas classes diferentes de marcadores moleculares). Um mapa de ligação que ilustra um conjunto bem-espaçado destes marcadores, com base nas análises da frequência de recombinantes do tipo descrito neste capítulo, está no centro da ilustração. As distâncias de mapa estão demonstradas em centimorgans (cM). Com um comprimento total de 356 cM, o cromossomo 1 é o mais longo dos cromossomos humanos. Também foram localizados alguns marcadores no mapa citogenético do cromossomo 1 (mapa à direita, denominado um idiograma), por meio da utilização das técnicas descritas posteriormente neste capítulo. A existência de marcadores em comum nos diferentes mapas genéticos possibilita que as localizações de outros genes e marcadores moleculares sejam estimadas em cada mapa. (Dados de B. R. Jasny *et al.*, *Science*, 30 de setembro de 1994.)

Em sua forma mais simples, o mapeamento do centrômero considera um *locus* gênico e indaga quão distante esse *locus* está do centrômero. O método tem por base o fato de que um padrão diferente de alelos aparecerá em uma tétrade linear ou uma óctade que tenha se originado de uma meiose com um *crossover* entre um gene e seu centrômero. Considere um cruzamento entre dois indivíduos, cada um deles apresentando um alelo diferente em um *locus* (digamos, $A \times a$). A lei da segregação igual de Mendel dita que, em uma óctade, sempre haverá quatro ascósporos do genótipo *A* e quatro do genótipo *a*, mas como eles estarão dispostos? Se não houve *crossover* na região entre *A*/*a* e o centrômero, haverá dois blocos adjacentes de quatro ascósporos na óctade linear (ver Figura 3.10, no Capítulo 3). Entretanto, se houve um *crossover* naquela região, haverá um de quatro padrões diferentes na óctade, cada um dos padrões demonstrando *blocos de dois alelos idênticos adjacentes*. Alguns dados de um cruzamento real de $A \times a$ estão demonstrados na tabela a seguir:

Óctades					
A	*a*	*A*	*a*	*A*	*a*
A	*a*	*A*	*a*	*A*	*a*

A	*a*	*a*	*A*	*a*	*A*
A	*a*	*a*	*A*	*a*	*A*
a	*A*	*A*	*a*	*a*	*A*
a	*A*	*A*	*a*	*a*	*A*
a	*A*	*a*	*A*	*A*	*a*
a	*A*	*a*	*A*	*A*	*a*
126	132	9	11	10	12

Total = 300

As primeiras duas colunas à esquerda são de meioses *sem crossover* na região entre o *locus A* e o centrômero. A letra M é utilizada para fazer referência a um tipo de segregação na meiose. Os padrões em relação às primeiras duas colunas são denominados **padrões de M_I**, ou **padrões de segregação de primeira divisão**, tendo em vista que os dois alelos diferentes segregam para dois núcleos-filhos na primeira divisão da meiose. As outras quatro colunas são todas de meiócitos *com* um *crossover*. Esses padrões são denominados **padrões de segregação de segunda divisão (M_{II})** tendo em vista que, como um resultado do *crossover* na região entre o centrômero e o *locus*, os alelos *A* e *a* ainda estão juntos nos núcleos ao término da primeira divisão da meiose (Figura 4.16). Não houve segregação na primeira divisão. Entretanto, a segunda divisão meiótica segrega os alelos *A* e *a* em núcleos separados. Os outros padrões são produzidos de modo semelhante; a diferença é que as cromátides se movimentam em direções diferentes na segunda divisão (Figura 4.17).

Você pode observar que a frequência de óctades com um padrão de M_{II} deve ser proporcional ao tamanho da região do centrômero até *A/a* e pode ser utilizada

como uma medida do tamanho daquela região. Em nosso exemplo, a frequência de M_{II} é 42/300 = 14%. Essa porcentagem significa que o *locus A/a* está a 14 u.m. do centrômero? A resposta é não, mas esse valor poderia ser utilizado para calcular o número de unidades de mapa. O valor de 14% é uma porcentagem da *meiose*, que não é o modo como as unidades de mapa são definidas. As unidades de mapa são definidas como a porcentagem de *cromátides* recombinantes oriundas da meiose. Tendo em vista que um *crossover* em qualquer meiose resulta em apenas 50% de cromátides recombinantes (quatro entre oito; ver Figura 4.17), é preciso dividir os 14% por 2 para converter a frequência de M_{II} (uma frequência de *meiose*) em unidades de mapa (uma frequência de *cromátides* recombinantes). Portanto, essa região tem obrigatoriamente um comprimento de 7 u.m. e essa medida pode ser introduzida no mapa daquele cromossomo.

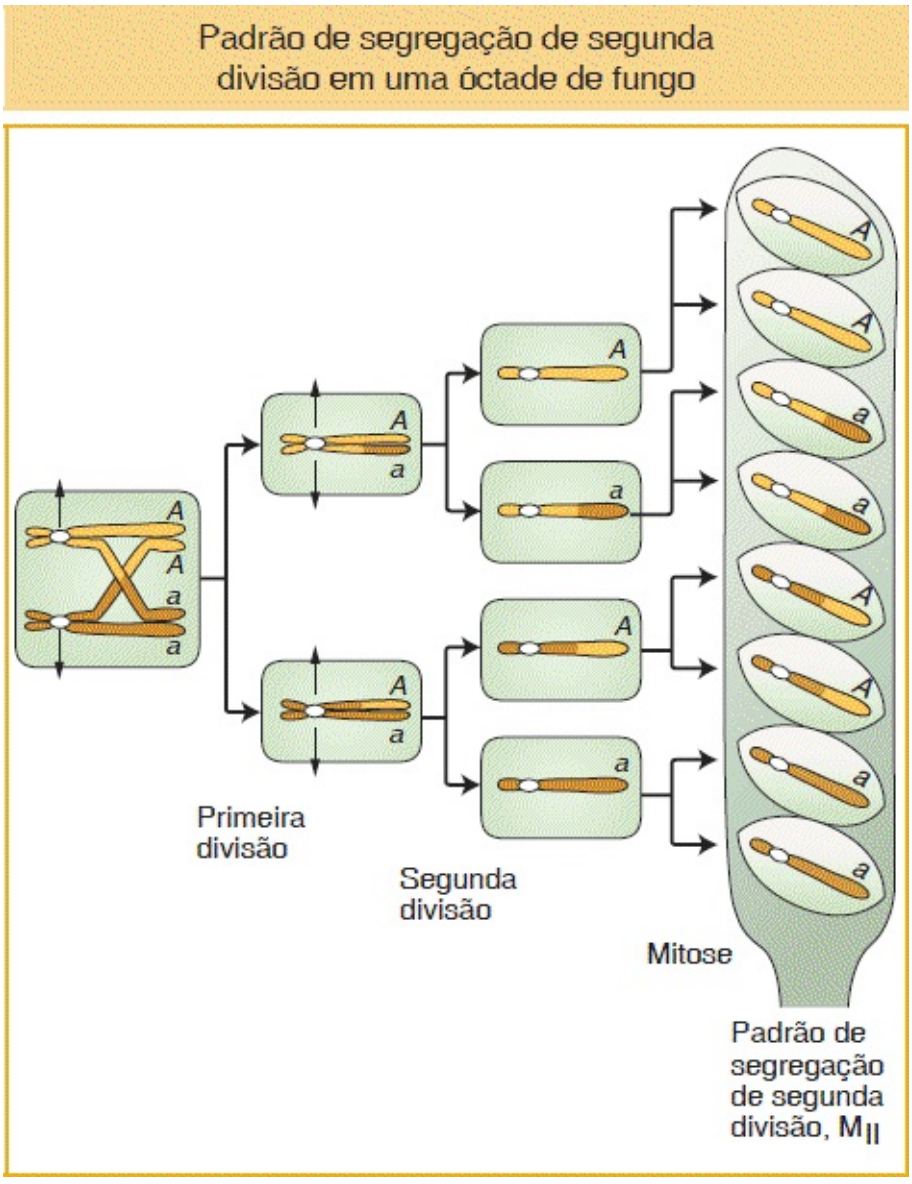


FIGURA 4.16 *A* e *a* segregam-se em núcleos separados na segunda divisão meiótica quando existe um *crossover* entre o centrômero e o *locus A*.

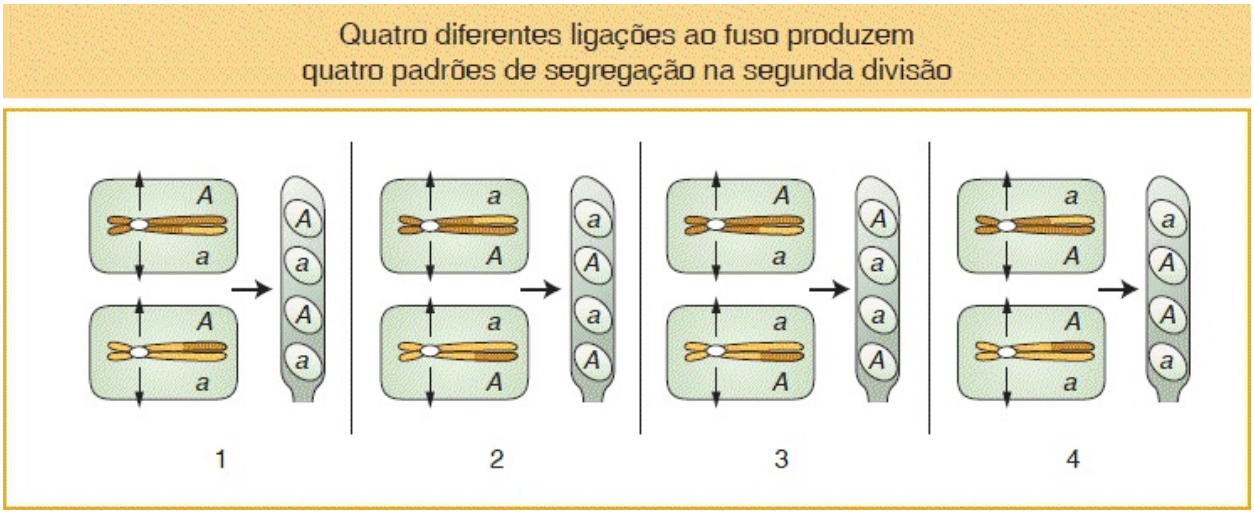


FIGURA 4.17 Na segunda divisão meiótica, os centrômeros se ligam ao fuso aleatoriamente, produzindo os quatro arranjos demonstrados. Os quatro arranjos são igualmente frequentes.

4.5 Utilização do teste do qui-quadrado para inferir ligação

O teste genético padrão em relação à ligação é um cruzamento-teste di-híbrido. Considere um cruzamento geral daquele tipo, no qual não se sabe se os genes estão ligados ou não:

$A/a \cdot B/b \times a/a \cdot b/b$

Se *não* existe ligação, ou seja, se os genes se distribuem de modo independente, observamos a partir das discussões neste capítulo e no Capítulo 3 que as proporções fenotípicas a seguir são esperadas na progênie:

A B	0,25
A b	0,25
a B	0,25
a b	0,25

Foi realizado um cruzamento desse tipo e os fenótipos a seguir foram obtidos em uma amostra da progênie de 200.

A B	60
A b	37
a B	41
a b	62

Claramente, existe um desvio da previsão de nenhuma ligação (que teria proporcionado os números de progênie de 50:50:50:50). Os resultados sugerem que o di-híbrido apresentava uma configuração cis de genes ligados, *A B/a b*, tendo em vista que as progênies *A B* e *a b* são a maioria. A frequência de recombinantes seria (37 + 41)/200 = 78/200 = 39%, ou 39 u.m.

Entretanto, sabemos que desvios ao acaso em virtude de erro de amostragem podem fornecer resultados que se assemelham àqueles produzidos pelos processos genéticos; portanto, precisamos do teste do χ^2 (pronunciado "qui-quadrado") para nos ajudar a calcular a probabilidade de um desvio ao acaso dessa magnitude de uma proporção de 1:1:1:1.

Primeiramente, examinemos as proporções de alelos em relação a ambos os *loci*. Essas são de 97:103 para *A:a* e 101:99 para *B:b*. Os referidos números estão próximos das proporções de alelos de 1:1 esperadas a partir da primeira lei de Mendel, de modo que proporções de alelos distorcidas não podem ser responsáveis pelos desvios consideravelmente grandes dos números esperados de 50:50:50:50.

É preciso aplicar a análise do χ^2 para testar uma hipótese de *nenhuma ligação*. Se a hipótese for rejeitada, podemos inferir a ligação. (Não podemos testar uma hipótese de ligação diretamente, tendo em vista que não temos como prever qual frequência de recombinantes deve ser testada.) O cálculo para o teste da ausência de ligação é como segue:

Observado (O)	Esperado (E)	$O - E$	$(O - E)^2$	$(O - E)^2/E$
60	50	10	100	2,00

37	50	—13	169	3,38
41	50	—9	81	1,62
62	50	12	144	2,88

$\chi^2 = \Sigma\ (O — E)^2/E$ para todas as classes = 9,88

Tendo em vista que existem quatro classes genotípicas, é preciso utilizar 4 — 1 = 3 graus de liberdade. Consultando a tabela do qui-quadrado no Capítulo 3, observamos que nossos valores de 9,88 e 3 gl fornecem um valor *p* de aproximadamente 0,025, ou 2,5%. Esse é inferior ao valor de corte padrão de 5% e, assim, podemos rejeitar a hipótese de nenhuma ligação. Portanto, chegamos à conclusão de que os genes muito provavelmente estão ligados, com distância de aproximadamente 39 u.m.

Observe, retrospectivamente, que foi muito importante assegurar que os alelos estavam segregando a 1:1 para evitar uma hipótese *composta* de proporções de alelos de 1:1 *e* nenhuma ligação. Se rejeitássemos uma referida hipótese composta, não saberíamos qual parte dela seria responsável pela rejeição.

4.6 Cômputo de crossovers múltiplos não visualizados

Na discussão sobre o cruzamento-teste de três pontos, algumas cromátides parentais (não recombinantes) resultaram de *crossovers duplos*. Esses *crossovers* inicialmente não podem ser contados na frequência de recombinantes, distorcendo os resultados. Essa situação leva à noção preocupante de que *todas* as distâncias de mapa com base na frequência de recombinantes podem ser subestimativas das distâncias físicas, tendo em vista que podem ter ocorrido *crossovers* múltiplos não detectados, alguns produtos dos quais não seriam recombinantes. Foram desenhadas diversas abordagens matemáticas criativas para contornar o problema do *crossover* múltiplo. Observaremos dois métodos. Primeiramente, examinamos

um método originalmente trabalhado por J. B. S. Haldane nos primórdios da genética.

Função de mapeamento

A abordagem trabalhada por Haldane foi o planejamento de uma **função de mapeamento**, uma fórmula que relaciona um valor de frequência de recombinantes observado com uma distância de mapa corrigida em relação a *crossovers* múltiplos. A abordagem relaciona FR com o número médio de *crossovers*, *m*, que devem ter ocorrido naquele segmento cromossômico pela meiose e em seguida deduz qual distância de mapa esse valor *m deve* ter produzido.

Para encontrar a relação de FR com *m*, primeiramente devemos pensar a respeito dos desfechos das diversas possibilidades de *crossover*. Em qualquer região cromossômica, devemos esperar meioses com 0, 1, 2, 3, 4 ou mais *crossovers*. Surpreendentemente, a única classe que é realmente crucial é a classe zero. Para saber o motivo, considere o que segue. Um fato curioso, porém não intuitivo, é que *qualquer quantidade de crossovers* produz uma frequência de 50% de recombinantes *naquelas meioses*. A Figura 4.18 comprova essa declaração em relação a *crossovers* únicos e duplos como exemplos, mas isso é verdadeiro em relação a qualquer número de *crossovers*. Portanto, o verdadeiro determinante da FR é o tamanho relativo das classes sem *crossovers* (a classe zero) em comparação às classes com qualquer número não zero de *crossovers*.

Agora a tarefa é calcular o tamanho da classe zero. A ocorrência de *crossovers* em uma região cromossômica específica é bem descrita por meio de uma distribuição estatística denominada **distribuição de Poisson**. A fórmula de Poisson em geral descreve a distribuição dos "sucessos" em amostras quando a probabilidade média de sucessos é baixa. Um exemplo ilustrativo é mergulhar uma pequena rede em um lago de peixes: a maior parte das imersões não pegará peixes, uma proporção menor pegará um peixe, uma proporção ainda menor pegará dois e assim por diante. Essa analogia pode ser aplicada diretamente a uma região cromossômica, que apresentará 0, 1, 2 *crossovers* de "sucesso" em diferentes meioses e assim por diante. A fórmula de Poisson, fornecida aqui, nos

informará a proporção das classes com diferentes números de *crossovers*:

$f_i = (e^{-m} m^i)/i!$

Qualquer número de *crossovers* fornece 50% de recombinantes

Nenhum *crossover*

RF = $\frac{0}{4}$ = **0%**

Um *crossover*

(Pode ser entre qualquer par não irmão.)

RF = $\frac{2}{4}$ = **50%**

Dois *crossovers*

(Manter um *crossover* constante e variar a posição do segundo produz quatro meioses de dois *crossovers* igualmente frequentes.)

RF = $\frac{0}{4}$ = **0%** — *Crossover* duplo de dois filamentos

RF = $\frac{2}{4}$ = **50%** — *Crossover* duplo de três filamentos

RF = $\frac{2}{4}$ = **50%** — *Crossover* duplo de três filamentos

RF = $\frac{4}{4}$ = **100%** — *Crossover* duplo de quatro filamentos

FR média de dois *crossovers* = $\frac{8}{16}$ = **50%**

FIGURA 4.18 Demonstração de que a FR média é de 50% em relação a meioses nas quais o número de *crossovers* não é zero. As cromátides recombinantes são marrons. *Crossovers* duplos de dois filamentos produzem todos os tipos parentais. Assim, todas as cromátides são laranja. Observe que todos os *crossovers* são entre cromátides não irmãs. Tente você mesmo a classe de *crossover* triplo.

Os termos na fórmula apresentam os significados a seguir:

e = A base dos logaritmos naturais (aproximadamente 2,7)
m = O número médio de sucessos em um tamanho definido de amostra
i = O número real de sucessos em uma amostra desse tamanho
f_i = A frequência das amostras que contém i sucessos
! = O símbolo de fatorial (p. ex., 5! = 5 × 4 × 3 × 2 × 1)

A distribuição de Poisson nos informa que a frequência da classe $i = 0$ (a classe-chave) é:

$$e^{-m}\frac{m^0}{0!}$$

Tendo em vista que m^0 e 0! são iguais a 1, a fórmula é reduzida a e^{-m}.

Agora podemos escrever uma função que relaciona FR com *m*. A frequência da classe com qualquer número não zero de *crossovers* será $1 - e^{-m}$, e, nessas meioses, 50% (1/2) dos produtos serão recombinantes. Assim:

$$FR = \tfrac{1}{2}(1 - e^{-m})$$

e essa fórmula é a função do mapeamento que estávamos buscando.

Observemos um exemplo no qual FR é convertida em uma distância de mapa corrigida em relação a *crossovers* múltiplos. Presuma que, em um cruzamento-teste, obtemos um valor de FR de 27,5% (0,275). Substituindo esse valor na função, podemos calcular *m*:

$$0{,}275 = \tfrac{1}{2}(1 - e^{-m})$$

Assim:

$e^{-m} = 1 — (2 \times 0{,}275) = 0{,}45$

Ao utilizar uma calculadora para encontrar o logaritmo natural (ln) de 0,45, podemos deduzir que m = 0,8. Ou seja, em média, existe 0,8 *crossover* por meiose naquela região cromossômica.

A etapa final é converter essa medida da frequência de *crossover* para fornecer uma distância de mapa "corrigida". Tudo o que precisamos fazer para a conversão em unidades de mapa corrigidas é multiplicar a frequência média de *crossover* calculada por 50, tendo em vista que, em média, um *crossover* produz uma frequência de recombinantes de 50%. Portanto, no exemplo numérico antecedente, o valor m de 0,8 pode ser convertido em uma fração recombinante corrigida de 0,8 × 50 = 40 u.m. corrigidas. Observamos que, de fato, esse valor é substancialmente maior do que as 27,5 u.m. que teríamos deduzido a partir da FR observada.

Observe que a função de mapeamento explica de modo organizado porque o valor máximo da FR para genes ligados é de 50%. Na medida em que m se torna muito grande, e^{-m} tende a zero e a FR tende a 1/2, ou 50%.

Fórmula de Perkins

Para os fungos e outros organismos que produzem tétrades, existe outro modo de compensar os *crossovers* múltiplos — especificamente, *crossovers* duplos (o tipo mais comum esperado). Na análise de tétrades de "di-híbridos" em geral, apenas três tipos de tétrades são possíveis, quando classificadas com base na presença de genótipos parentais e recombinantes nos produtos. A classificação das tétrades tem por base a existência de dois (ditipo) ou quatro (tetratipo) genótipos. Nos ditipos, existem duas classes: parental (que demonstra dois genótipos parentais) e não parental (que demonstra dois genótipos não parentais). A partir de um cruzamento *AB* × *ab*, elas são:

Ditipo parental (DP)	Tetratipo (T)	Ditipo não parental (DNP)

$A \cdot B$	$A \cdot B$	$A \cdot b$
$A \cdot B$	$A \cdot b$	$A \cdot b$
$a \cdot b$	$a \cdot B$	$a \cdot B$
$a \cdot b$	$a \cdot b$	$a \cdot B$

Os genótipos recombinantes estão demonstrados em vermelho. Se os genes estiverem ligados, uma abordagem simples para mapear sua distância pode ser a utilização da fórmula a seguir:

Distância de mapa = FR = 100 (DNP + $\frac{1}{2}$ T)

tendo em vista que essa fórmula fornece a porcentagem de todos os recombinantes. Entretanto, na década de 1960, David Perkins desenvolveu uma fórmula que compensa os efeitos de *crossovers* duplos. A fórmula de Perkins, portanto, proporciona uma estimativa mais exata da distância de mapa:

Distância de mapa corrigida = 50 (T + 6 DNP)

Não desmembraremos mais essa fórmula, a não ser para dizer que ela tem por base os totais das classes de DP, T e DNP esperadas a partir de meioses com 0, 1 e 2 *crossovers* (ela presume que números mais altos são cada vez mais raros). Observemos um exemplo da sua utilização. Em nosso cruzamento hipotético de *A B* × *a b*, as frequências observadas das classes de tétrades são DP 0,56, T 0,41 e DNP 0,03. Ao utilizar a fórmula de Perkins, observamos que a distância de mapa corrigida entre os *loci a* e *b* é:

50 [0,41 + (6 × 0,03)] = 50 (0,59) = 29,5 u.m.

Comparemos esse valor com o valor não corrigido obtido diretamente da FR. Ao utilizar os mesmos dados, encontramos

Distância de mapa não corrigida	=	100 ($\frac{1}{2}$ T + DNP)
	=	100 (0,205 + 0,03)
	=	23,5 u.m.

Essa distância é 6 u.m. inferior à estimativa que obtivemos por meio da utilização da fórmula de Perkins, tendo em vista que não a corrigimos em relação a *crossovers* duplos.

Como um comentário à parte, quais valores de DP, DNP e T são esperados quando lidamos com genes *não ligados*? Os tamanhos das classes de DP e DNP serão iguais, como um resultado da distribuição independente. A classe T pode ser produzida apenas a partir de um *crossover* entre qualquer um dos dois *loci* e seus respectivos centrômeros e, portanto, o tamanho da classe T dependerá do tamanho total das duas regiões que estão localizadas entre o *locus* e o centrômero. Entretanto, a fórmula $\frac{1}{2}$ T + DNP sempre deve produzir 0,50, refletindo a distribuição independente.

> CONCEITO-CHAVE A tendência inerente de *crossovers* múltiplos de levar a uma subestimativa da distância de mapa pode ser contornada por meio da utilização das funções de mapa (em qualquer organismo) e pela fórmula de Perkins (em organismos que produzem tétrades, tais como fungos).

4.7 Utilização de mapas com base em recombinação em conjunto com mapas físicos

Os mapas de recombinação têm sido o principal tópico deste capítulo. Eles demonstram os *loci* de genes em relação aos quais foram observados alelos mutantes (e seus fenótipos mutantes). As posições desses *loci* em um mapa são determinadas com base na frequência de recombinantes na meiose. Presume-se que a frequência de recombinantes seja proporcional à distância que separa dois *loci* no cromossomo; portanto, a frequência de recombinantes se torna a unidade

de mapeamento. O referido mapeamento com base na recombinação de genes com fenótipos mutantes conhecidos vem sendo realizado por quase um século. Observamos como os sítios de heterozigosidade molecular (não associados a fenótipos mutantes) também podem ser incorporados a esses mapas de recombinação. Assim como qualquer sítio heterozigoto, esses marcadores moleculares são mapeados por meio de recombinação e, em seguida, são utilizados para navegar em direção a um gene de interesse biológico. Realizamos a presunção perfeitamente razoável de que um mapa de recombinação representa o arranjo de genes nos cromossomos, mas, conforme declarado anteriormente, esses mapas na realidade são constructos hipotéticos. Contrariamente, os mapas físicos são o mais próximo do mapa genômico real que a ciência consegue obter.

O tópico dos **mapas físicos** será examinado mais de perto no Capítulo 14, mas podemos indicá-lo aqui. Um mapa físico é simplesmente um mapa do DNA genômico real, uma sequência de nucleotídios muito longa, que demonstra onde os genes estão, a sua sequência, quão grandes eles são, o que se encontra entre eles e outros marcos de interesse. As unidades de distância em um mapa físico são os números de bases do DNA; por conveniência, o quilobase é a unidade preferida. A sequência completa de uma molécula de DNA é obtida por meio do sequenciamento de grandes números de pequenos fragmentos genômicos e em seguida da sua montagem em uma sequência integral. A sequência em seguida é examinada por um computador programado para procurar por segmentos semelhantes a genes reconhecidos por meio de sequências de bases características, incluindo sequências de sinal conhecidas em relação ao início e ao término da transcrição. Quando o programa de computador encontra um gene, ele compara sua sequência com a base de dados pública de outros genes sequenciados em relação aos quais foram descobertas as funções em outros organismos. Em muitos casos, ocorre um “achado”; em outras palavras, a sequência se assemelha de modo próximo àquela de um gene de função conhecida em outra espécie. Nos referidos casos, as funções dos dois genes também podem ser semelhantes. A similaridade de sequência (com frequência próxima dos 100%) é explicada pela herança do gene a partir de algum ancestral comum e pela conservação geral de sequências funcionais durante o período evolutivo. Outros

genes descobertos pelo computador não demonstram similaridade de sequência com qualquer gene de função conhecida. Portanto, eles podem ser considerados "genes em busca de uma função". É claro que, na realidade, é o pesquisador, não o gene, que realiza a procura e que deve encontrar a função. O sequenciamento de diferentes indivíduos, membros de uma população, também pode produzir sítios de heterozigosidade molecular, os quais, assim como o fazem nos mapas de recombinação, atuam como marcadores de orientação no mapa físico.

Tendo em vista que atualmente estão disponíveis mapas físicos em relação à maior parte dos principais organismos-modelo genéticos, realmente existe necessidade de mapas de recombinação? Eles poderiam ser considerados obsoletos? A resposta é que ambos os mapas são utilizados em conjunto para "triangular" a determinação da função gênica, um princípio ilustrado anteriormente pelos mapas de Londres. A abordagem geral está ilustrada na Figura 4.19, que demonstra um mapa físico e um mapa de recombinação da mesma região de um genoma. Ambos os mapas contêm genes e marcadores moleculares. Na parte inferior da Figura 4.19, observamos uma seção de um mapa com base na recombinação, com as posições dos genes em relação aos quais foram encontrados e mapeados fenótipos mutantes. Nem todos os genes naquele segmento estão incluídos. Em relação a alguns desses genes, pode ter sido descoberta uma função com base em estudos bioquímicos ou outros estudos de linhagens mutantes; os genes em relação às proteínas A e B são exemplos. O gene na parte intermediária é um "gene de interesse" que um pesquisador observou afetar o aspecto do desenvolvimento que está sendo estudado. Para determinar a sua função, o mapa físico pode ser útil. Os genes no mapa físico que estão na região geral do gene de interesse no mapa de recombinação se tornam *genes candidatos*, qualquer um dos quais pode ser o gene de interesse. São necessários estudos adicionais para estreitar a escolha até um. Se aquele caso único for um gene cuja função é conhecida em outros organismos, então é sugerida uma função para o gene de interesse. Desse modo, o fenótipo mapeado no mapa de recombinação pode ser associado a uma função deduzida a partir do mapa físico. Marcadores moleculares em ambos os mapas (não demonstrados na Figura 4.19) podem ser alinhados para auxiliar no processo de seleção. Portanto, observamos

que ambos os mapas contêm elementos da função: o mapa físico demonstra a possível ação de um gene no nível celular, enquanto o mapa de recombinação contém informação relacionada ao efeito do gene no nível fenotípico. Em algum estágio, os dois precisam ser combinados para a compreensão da contribuição do gene para o desenvolvimento do organismo.

Existem diversas outras técnicas de mapeamento genético, algumas das quais encontraremos nos Capítulos 5, 18 e 19.

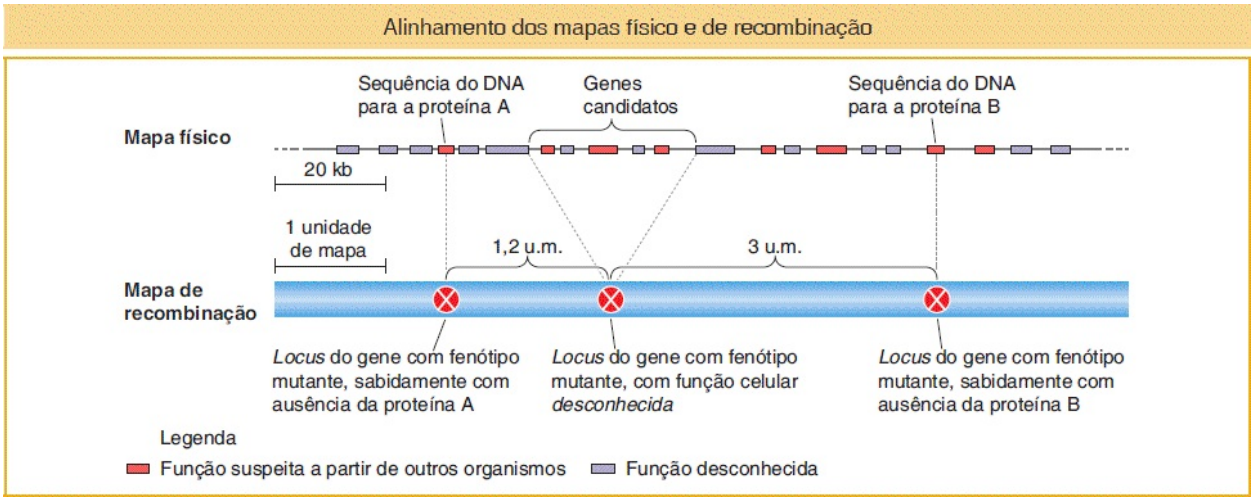


FIGURA 4.19 Comparação das posições relativas em mapas físicos e de recombinação pode conectar o fenótipo a uma função gênica desconhecida.

> CONCEITO-CHAVE A união dos mapas físico e de recombinação permite atribuir uma função bioquímica a um gene identificado por meio de seu fenótipo mutante.

4.8 Mecanismo molecular de crossover

Neste capítulo, analisamos as consequências genéticas do processo citologicamente visível de *crossover* sem nos preocuparmos a respeito do mecanismo do *crossing over*. Entretanto, o *crossing over*, por si próprio, é um processo molecular extraordinário: como duas grandes moléculas de DNA

espiraladas podem trocar segmentos com tamanha precisão que nenhum nucleotídio é perdido ou ganho?

Estudos em óctades de fungos forneceram um indício. Embora a maior parte das óctades demonstre a esperada segregação de alelos de 4:4 tais como 4*A*:4*a*, algumas óctades raras demonstram proporções aberrantes. Existem diversos tipos, mas como um exemplo, utilizaremos óctades de 5:3 (sejam 5*A*:3*a* ou 5*a*:3*A*). Dois aspectos são peculiares dessa proporção. Primeiramente, existem muitos esporos de um alelo e muito poucos do outro. Em segundo lugar, existe um *par de esporos-irmãos não idêntico*. Normalmente, a replicação pós-meiótica origina pares de esporos-irmãos idênticos, como segue: a tétrade *A A a a* se torna:

A-*A*-*A*-*A a*-*a*-*a*-*a*

(os hifens demonstram esporos-irmãos). Contrariamente, uma óctade aberrante de 5*A*:3*a* deve ser:

A-*A A*-*A* ***A*-*a*** *a*-*a*

Em outras palavras, existe um par de esporos-irmãos não idêntico (em negrito).

A observação de um par de esporos-irmãos não idêntico sugere que o DNA de um dos quatro homólogos meióticos finais contém **DNA heterodúplex**. O DNA heterodúplex é o DNA no qual existe um par de nucleotídios incorretamente pareados no gene em estudo. A lógica é como segue. Se em um cruzamento de *A* × *a*, um alelo (*A*) for G:C e o outro alelo (*a*) for A:T, os dois alelos normalmente se replicam fielmente. Entretanto, um heterodúplex, que se forma apenas raramente, seria um par de nucleotídios incorretamente pareados, tal como G:T ou A:C (efetivamente uma molécula de DNA que contém ambas as informações de *A* e *a*). Observe que um heterodúplex envolve apenas a posição de um nucleotídio: o segmento de DNA adjacente pode ser como segue, onde o sítio do heterodúplex está demonstrado em negrito:

GCTAAT**G**TTATTAG
CGATTA**T**AATAATC

Na replicação para formar uma óctade, um heterodúplex G:T seria separado e replicado fielmente, com a ligação de G com C e a ligação de A com T. O resultado seria um par de esporos não idênticos de G:C (alelo *A*) e A:T (alelo *a*).

Observou-se que esporos-irmãos não idênticos (e em geral óctades aberrantes) estão estatisticamente correlacionados com o *crossing over* na região do gene em questão, fornecendo uma indicação importante de que o *crossing over* poderia basear-se na formação de DNA heterodúplex.

No modelo atualmente aceito (acompanhe-o na Figura 4.20), ambos o DNA heterodúplex e um *crossover* são produzidos por uma **quebra do filamento duplo** no DNA de uma das cromátides que participam no *crossover*. Vejamos como isto ocorre. Estudos moleculares demonstram que as extremidades do DNA quebradas promoverão a recombinação entre diferentes cromátides. Na etapa 1, ambos os filamentos de uma cromátide quebram no mesmo local. A partir da quebra, o DNA é erodido na extremidade 5' de cada filamento quebrado, deixando ambas as extremidades 3' em filamento simples (etapa 2). Um dos filamentos simples "invade" o DNA da outra cromátide participante; ou seja, ele entra no centro da hélice e dos pares de bases com a sua sequência homóloga (etapa 3), deslocando o outro filamento. Em seguida, a ponta do filamento invasor utiliza a sequência adjacente como um molde para a nova polimerização, que prossegue forçando os dois filamentos residentes da hélice a se separarem (etapa 4). Ligações de hidrogênio formam-se entre a alça de filamento simples deslocada e o outro filamento simples (o filamento azul na figura). Se a invasão e o deslocamento do filamento abrangerem um sítio de heterozigosidade (tal como *A/a*), então é formada uma região de DNA heterodúplex. A replicação também ocorre a partir da outra extremidade do filamento simples para preencher o espaço deixado pelo filamento invasor (também demonstrado no filamento azul superior na etapa 4 da Figura 4.20). As extremidades replicadas são seladas e o resultado líquido é uma estrutura estranha, com duas junções de filamento simples denominadas *junções de Holliday*, em homenagem ao seu proponente original, Robin Holliday. Essas junções são sítios potenciais de quebra e reunião de filamentos simples; os dois referidos eventos, demonstrados pelos dardos na figura, levam então a um *crossover* completo de filamento duplo (etapa 5).

O *crossing over* cria DNA heterodúplex

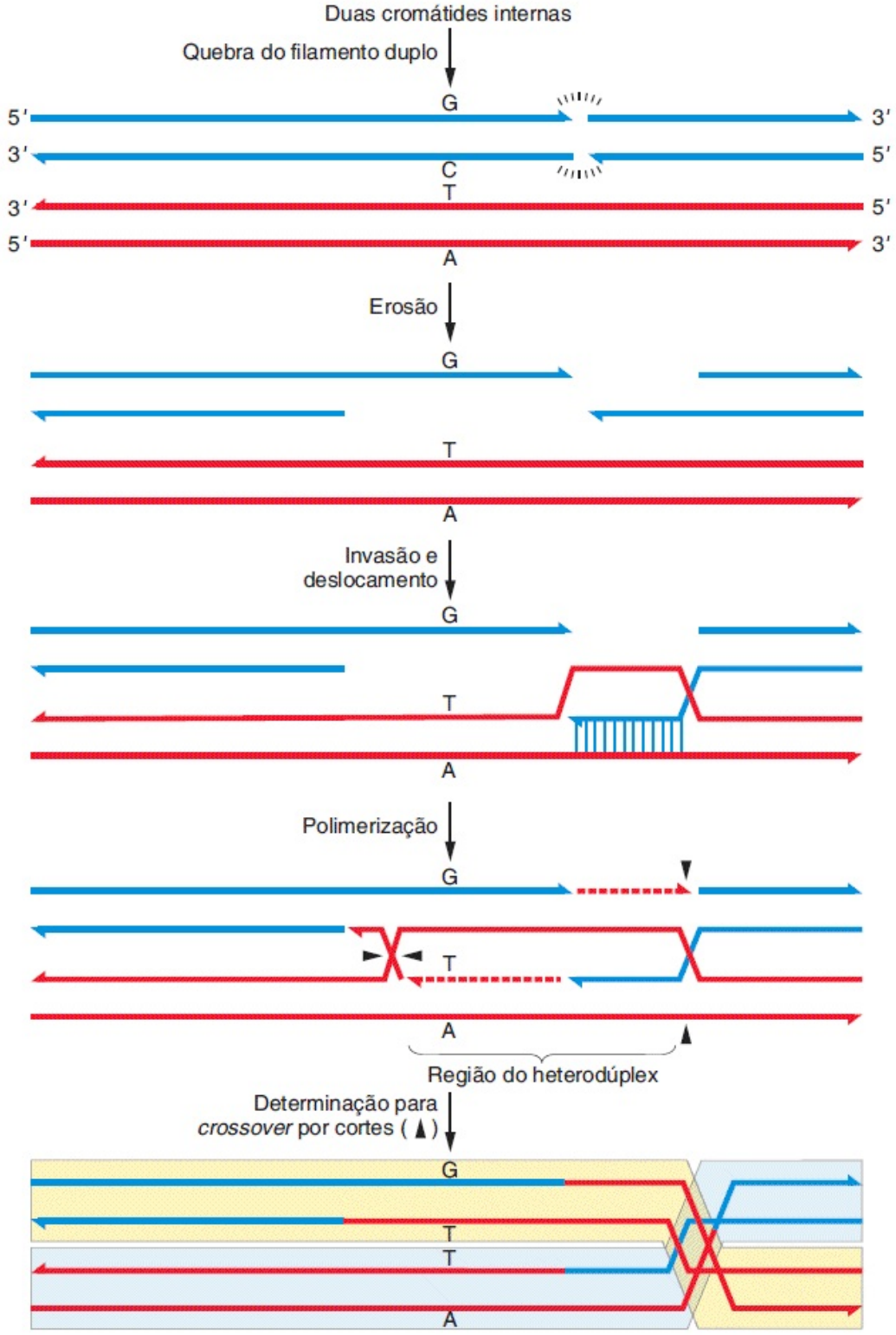


FIGURA 4.20 Um modelo molecular de *crossing over*. Apenas as duas cromátides (*azul* e *vermelha*) que participam no *crossover* estão demonstradas. O sentido do filamento 3' para 5' consta de ambos para maior clareza. As cromátides diferem em um sítio, GC, em um alelo (talvez o alelo *A*) e AT no outro (talvez *a*). Apenas o desfecho com DNA heterodúplex incorretamente pareado e um *crossover* estão demonstrados. Os produtos finais do *crossover* estão sombreados em *amarelo* e *azul*.

Observe que quando o filamento invasor utiliza o DNA invadido como molde para a replicação, isso resulta automaticamente em uma cópia extra da sequência invadida à custa da sequência invasora, explicando, assim, o afastamento da proporção esperada de 4:4.

Esse mesmo tipo de recombinação ocorre em muitos locais cromossômicos diferentes, nos quais a invasão e o deslocamento do filamento *não* abrangem um sítio heterozigoto mutante. Aqui, seria formado um DNA que é heterodúplex no sentido em que é composto por filamentos de cada cromátide participante, mas não haveria um par de nucleotídios incorretamente pareados e a óctade resultante conteria apenas pares de esporos idênticos. Aquelas ocasiões raras nas quais a invasão e a polimerização *abrangem* um sítio heterozigoto são simplesmente casos de sorte, que forneceram a indicação em relação ao mecanismo de *crossing over*.

> CONCEITO-CHAVE Um *crossover* é iniciado por meio de uma quebra do filamento duplo no DNA de uma cromátide na meiose. Segue-se uma série de eventos moleculares, que finalmente produzem o *crossover* de moléculas de DNA. (Além disso, se o sítio do *crossover* estiver próximo de um sítio de heterozigosidade do DNA na meiose, podem ser produzidas proporções não mendelianas aberrantes de alelos em relação ao sítio heterozigoto.)

RESUMO

Em um cruzamento-teste di-híbrido em *Drosophila*, Thomas Hunt Morgan encontrou um desvio da lei de Mendel sobre a distribuição independente. Ele postulou que os dois genes estavam localizados no mesmo par de cromossomos homólogos. Essa relação é denominada ligação.

A ligação explica por que as combinações gênicas parentais permanecem juntas, mas não como surgem as combinações recombinantes (não parentais). Morgan postulou que, na meiose, pode haver uma troca física de partes dos cromossomos por meio de um processo atualmente denominado *crossover*. Um resultado da quebra física e da reunião de partes dos cromossomos, o *crossover* ocorre no estágio de quatro cromátides da meiose. Portanto, existem dois tipos de recombinação meiótica. A recombinação por meio da distribuição independente mendeliana resulta em uma frequência de recombinantes de 50%. O *crossover* resulta em uma frequência de recombinantes em geral inferior a 50%.

Na medida em que Morgan estudava mais genes ligados, ele descobriu muitos valores diferentes em relação à frequência de recombinantes e ponderou se esses valores correspondiam às distâncias reais entre genes em um cromossomo. Alfred Sturtevant, um aluno de Morgan, desenvolveu um método para determinar a distância entre genes em um mapa de ligação, com base na FR. O modo mais fácil de medir a FR é com um cruzamento-teste de um di-híbrido ou tri-híbrido. Os valores de FR calculados como porcentagens podem ser utilizados como unidades de mapa para construir um mapa cromossômico que demonstre os *loci* dos genes analisados. Em fungos ascomicetos, os centrômeros também podem ser localizados no mapa por meio da medição das frequências de segregação da segunda divisão.

Os polimorfismos de nucleotídio único (SNP) são diferenças de um único nucleotídio nas sequências do DNA. Os polimorfismos de comprimento de sequência simples (SSLP) são diferenças no número de unidades repetidas. Os SNP e SSLP podem ser utilizados como marcadores moleculares para o mapeamento de genes.

Embora o teste básico em relação à ligação seja o desvio da distribuição independente, um referido desvio pode não ser óbvio em um cruzamento-teste e é necessário um teste estatístico. O teste do χ^2, que nos informa com que frequência as observações se desviam das expectativas puramente ao acaso, é particularmente útil para determinar se os *loci* estão ligados.

Alguns *crossovers* múltiplos podem resultar em cromátides não recombinantes, levando a uma subestimativa da distância de mapa com base na FR. A função de

mapeamento, aplicável em qualquer organismo, corrige em relação a essa tendência. A fórmula de Perkins apresenta a mesma utilização na análise de tétrades de fungos.

Em geral na genética, o mapa de *loci* com base na recombinação, que confere os fenótipos mutantes, é utilizado em conjunto com um mapa físico, como a sequência completa do DNA, que demonstra todas as sequências semelhantes a genes. O conhecimento da posição do gene em ambos os mapas possibilita associar a função celular ao efeito do gene no fenótipo.

Acredita-se que o mecanismo de *crossover* tenha início com uma quebra do filamento duplo em uma cromátide participante. A erosão deixa as extremidades em filamento simples. Um filamento simples invade a dupla-hélice da outra cromátide participante, levando à formação do DNA heterodúplex. Os espaços são preenchidos por polimerização. A resolução molecular dessa estrutura se torna um *crossover* de filamento duplo total no nível do DNA.

TERMOS-CHAVE

centimorgan (cM)
coeficiente de coincidência (c.d.c.)
conformação cis
conformação trans
crossing over
cruzamento de três fatores
cruzamento-teste de três pontos
distribuição de Poisson
DNA heterodúplex
frequência de recombinantes (FR)
função de mapeamento
genes ligados
impressões digitais do DNA
interferência
locus

mapa cromossômico
mapa de ligação
mapa de recombinação
mapa físico
marcador microssatélite
marcador minissatélite
marcador molecular
óctade
padrão de segregação de primeira divisão (padrão de M_I)
padrão de segregação de segunda divisão (padrão de M_{II})
polimorfismo de comprimento de fragmento de restrição (RFLP)
polimorfismo de comprimento de sequência simples (SSLP)
polimorfismo de nucleotídio único (SNP)
produto de *crossover*
quebra do filamento duplo
repetições em tandem de número variável (VNTR)
unidade de mapa genético (u.m.)

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Um heredograma humano demonstra pessoas afetadas pela rara síndrome unha-patela (unhas e rótulas deformadas) e fornece o genótipo do grupo sanguíneo AB0 de cada pessoa. Ambos os *loci* em questão são autossômicos. Estude o heredograma a seguir.

a. A síndrome unha-patela é um fenótipo dominante ou recessivo? Forneça os motivos para amparar sua resposta.

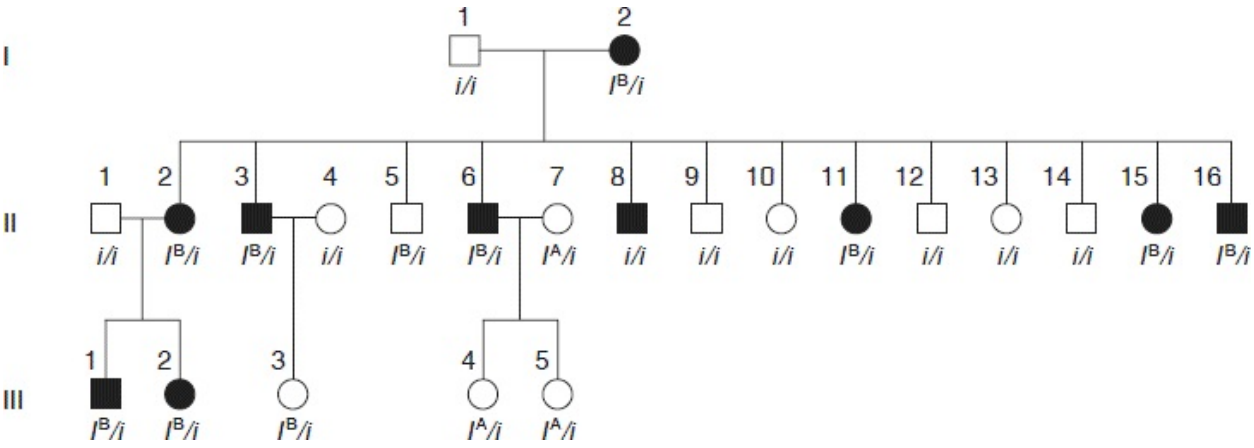
b. Existe evidência de ligação entre o gene relacionado à unha-patela e o gene do tipo sanguíneo AB0, conforme analisado a partir desse heredograma? Por que sim ou por que não?

c. Se houver evidência de ligação, em seguida desenhe os alelos nos homólogos relevantes dos avós. Se não houver evidência de ligação, desenhe os alelos em dois pares homólogos.

d. De acordo com o seu modelo, quais descendentes da geração II são recombinantes?

e. Qual é a melhor estimativa de FR?

f. Se o homem III-1 se casar com uma mulher normal de tipo sanguíneo O, qual é a probabilidade de que seu primeiro filho seja do tipo sanguíneo B com síndrome unha-patela?



Solução

a. A síndrome unha-patela é mais provavelmente dominante. É-nos informado que ela é uma anormalidade rara e, assim, as pessoas não afetadas que se casam com pessoas da família provavelmente não carreiam um alelo presumidamente recessivo para a síndrome unha-patela. Deixe que *N* seja o alelo causador. Então todas as pessoas com a síndrome são heterozigotas *N/n*, tendo em vista que todas (provavelmente incluindo a avó) resultam de cruzamentos com pessoas normais *n/n*. Observe que a síndrome aparece em todas as três gerações — outra indicação de herança dominante.

b. Existe evidência de ligação. Observe que a maior parte das pessoas afetadas — aquelas que carregam o alelo *N* — também carregam o alelo I^B; mais provavelmente, esses alelos estão ligados no mesmo cromossomo.

c. $$\frac{n \quad i}{n \quad i} \times \frac{N \quad I^B}{n \quad i}$$

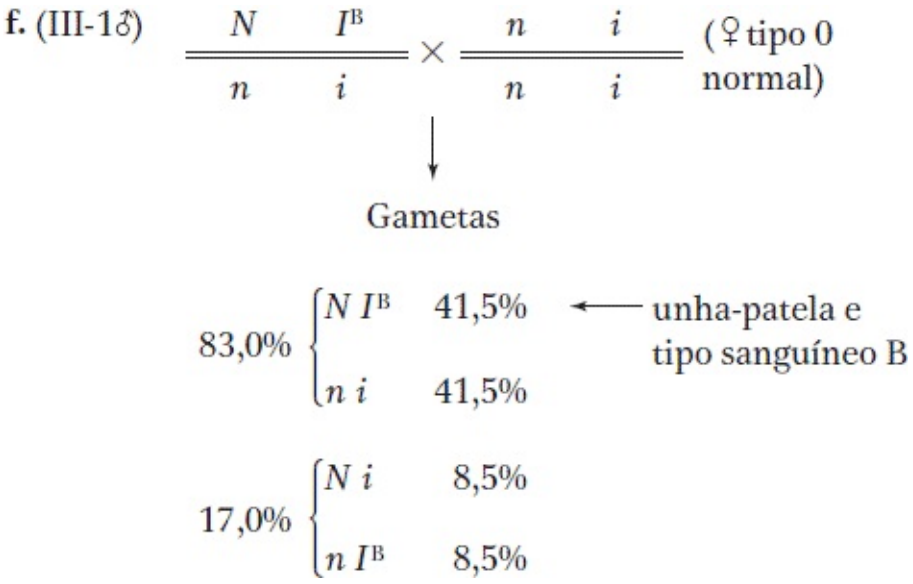
(A avó deve carregar ambos os alelos recessivos para produzir descendência de genótipo *i/i* e *n/n*.)

d. Observe que o cruzamento dos avós é equivalente a um cruzamento-teste; assim, os recombinantes na geração II são:

II-5: *n* I^B/*n i* e II-8: *N i*/*n i*

enquanto todos os outros são não recombinantes, sendo *N* I^B/*n i* ou *n i*/*n i*.

e. Observe que o cruzamento dos avós e os primeiros dois cruzamentos na geração II são idênticos e são cruzamentos-teste. Três da progênie total de 16 são recombinantes (II-5, II-8 e III-3). O cruzamento de II-6 com II-7 não é um cruzamento-teste, mas pode-se deduzir que os cromossomos doados de II-6 são não recombinantes. Portanto, FR = 3/18, que é 17%.



As duas classes parentais são sempre iguais e, assim, também são as duas classes recombinantes. Portanto, a probabilidade de que o primeiro filho apresente síndrome unha-patela e grupo sanguíneo tipo B é de 41,5%.

Problema resolvido 2. O alelo *b* proporciona às moscas *Drosophila* um corpo preto e b^+ proporciona uma cor marrom, o fenótipo do tipo selvagem. O alelo *wx* de um gene separado proporciona asas cerosas e wx^+ proporciona não cerosas, o fenótipo do tipo selvagem. O alelo *cn* de um terceiro gene proporciona olhos cor de cinábrio e cn^+ proporciona olhos vermelhos, o fenótipo do tipo selvagem. Uma fêmea heterozigota em relação a esses três genes é submetida a um cruzamento-

teste e uma progênie de 1.000 é classificada como segue: 5 do tipo selvagem; 6 pretas, cerosas, cinábrio; 69 cerosas, cinábrio; 67 pretas; 382 cinábrio; 379 pretas, cerosas; 48 cerosas; e 44 pretas, cinábrio. Observe que um grupo da progênie pode ser especificado por meio da listagem apenas dos fenótipos mutantes.

a. Explique esses números.

b. Desenhe os alelos em suas posições adequadas nos cromossomos do heterozigoto triplo.

c. Se apropriado, de acordo com a sua explicação, calcule a interferência.

Solução

a. Um conselho geral é ser metódico. Aqui, é uma boa ideia escrever os genótipos que podem ser inferidos a partir dos fenótipos. O cruzamento é um cruzamento-teste do tipo:

$$b^+/b \cdot wx^+/wx \cdot cn^+/cn \times b/b \cdot wx/wx \cdot cn/cn$$

Observe que existem pares distintos de classes de progênie em relação à frequência. Já podemos adivinhar que as duas classes maiores representam cromossomos parentais, que as duas classes de aproximadamente 68 representam *crossovers* únicos em uma região, que as duas classes de aproximadamente 45 representam *crossovers* únicos na outra região e que as duas classes de aproximadamente 5 representam *crossovers* duplos. Podemos escrever a progênie como classes derivadas dos gametas femininos, agrupados como segue:

$b^+ \cdot wx^+ \cdot cn$	382
$b \cdot wx \cdot cn^+$	379
$b^+ \cdot wx \cdot cn$	69
$b \cdot wx^+ \cdot cn^+$	67
$b^+ \cdot wx \cdot cn^+$	48
$b \cdot wx^+ \cdot cn$	44
$b \cdot wx \cdot cn$	6

$b^+ \cdot wx^+ \cdot cn^+$	5
	1.000

Listar as classes desse modo confirma que os pares de classes são, de fato, genótipos recíprocos que surgem de zero, um ou dois *crossovers*.

Primeiramente, tendo em vista que não conhecemos os genitores da fêmea heterozigota tripla, aparentemente não podemos aplicar a definição de recombinação, na qual os genótipos gaméticos são comparados aos dois genótipos parentais que formam uma mosca individual. Mas, com a reflexão, os únicos tipos parentais que fazem sentido em relação aos dados apresentados são $b^+/b^+ \cdot wx^+/wx^+ \cdot cn/cn$ e $b/b \cdot wx/wx \cdot cn^+/cn^+$, tendo em vista que esses tipos representam as classes gaméticas mais comuns.

Agora, podemos calcular as frequências de recombinantes. Em relação a *b*-*wx*:

$$\text{FR} = \frac{69+67+48+44}{1.000} = 22{,}8\%$$

em relação a *b*-*cn*:

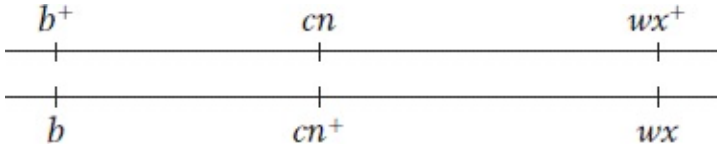
$$\text{FR} = \frac{48+44+6+5}{1.000} = 10{,}3\%$$

e em relação a *wx*-*cn*:

$$\text{FR} = \frac{69+67+6+5}{1.000} = 14{,}7\%$$

Portanto, o mapa é:

b. Os cromossomos parentais no heterozigoto triplo são:



c. O número esperado de recombinantes duplos é 0,103 × 0,147 × 1.000 = 15,141. O número observado é 6 + 5 = 11 e, assim, a interferência pode ser calculada como:

I = 1 — (11/15,141) = 1 — 0,726 = 0,274 = 27,4%

Problema resolvido 3. É realizado um cruzamento entre uma linhagem haploide de *Neurospora* de genótipo *nic*$^+$ *ad* e outra linhagem haploide de genótipo *nic ad*$^+$. A partir desse cruzamento, um total de 1.000 ascos lineares é isolado e categorizado como na tabela a seguir. Mapeie os *loci ad* e *nic* em relação aos centrômeros e uns aos outros.

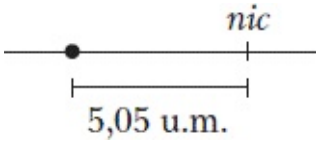
1	2	3	4	5	6	7
nic$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*$^+$
nic$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*$^+$
nic$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*· *ad*	*nic*· *ad*$^+$	*nic*· *ad*	*nic*· *ad*
nic$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*· *ad*	*nic*· *ad*$^+$	*nic*· *ad*	*nic*· *ad*
nic· *ad*$^+$	*nic*· *ad*	*nic*· *ad*$^+$	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*
nic· *ad*$^+$	*nic*· *ad*	*nic*· *ad*$^+$	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*	*nic*$^+$· *ad*$^+$	*nic*$^+$· *ad*
			nic·			

nic· *ad*$^+$	*nic*· *ad*	*nic*· *ad*	*ad*$^+$	*nic*· *ad*$^+$	*nic*· *ad*	*nic*· *ad*$^+$
nic· *ad*$^+$	*nic*· *ad*	*nic*· *ad*	*nic*· *ad*$^+$	*nic*· *ad*$^+$	*nic*· *ad*	*nic*· *ad*$^+$
808	1	90	5	90	1	5

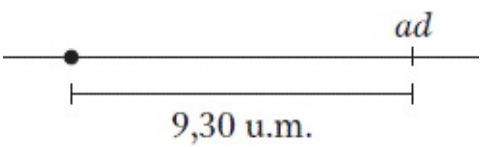
Solução

Em quais princípios podemos nos inspirar para solucionar este problema? É uma boa ideia iniciar realizando algo direto, que é calcular as duas distâncias do *locus* até o centrômero. Não sabemos se os *loci ad* e *nic* estão ligados, mas não precisamos saber. As frequências dos padrões de M_{II} em relação a cada *locus* fornecem a distância do *locus* até o centrômero. (Podemos nos preocupar se é o mesmo centrômero posteriormente.)

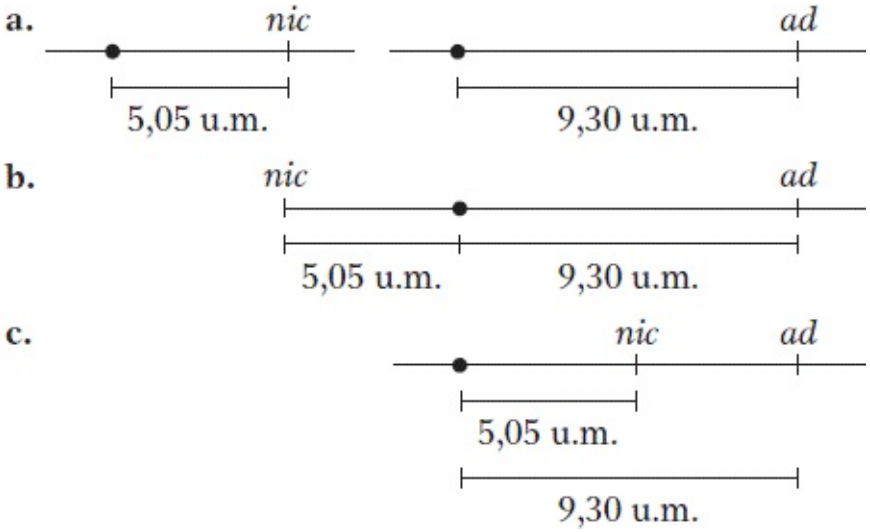
Relembre que um padrão de M_{II} é qualquer padrão que não é dois blocos de quatro. Iniciemos com a distância entre o *locus nic* e o centrômero. Tudo o que precisamos fazer é somar os ascos dos tipos 4, 5, 6 e 7, tendo em vista que todos eles são padrões de M_{II} em relação ao *locus nic*. O total é 5 + 90 + 1 + 5 = 101 de 1.000, ou 10,1%. Neste capítulo, vimos que para converter essa porcentagem em unidades de mapa, devemos dividir por 2, o que fornece 5,05 u.m.



Realizamos o mesmo em relação ao *locus ad*. Aqui, o total dos padrões de M_{II} é fornecido pelos tipos 3, 5, 6 e 7 e é 90 + 90 + 1 + 5 = 186 de 1.000, ou 18,6%, que é 9,3 u.m.

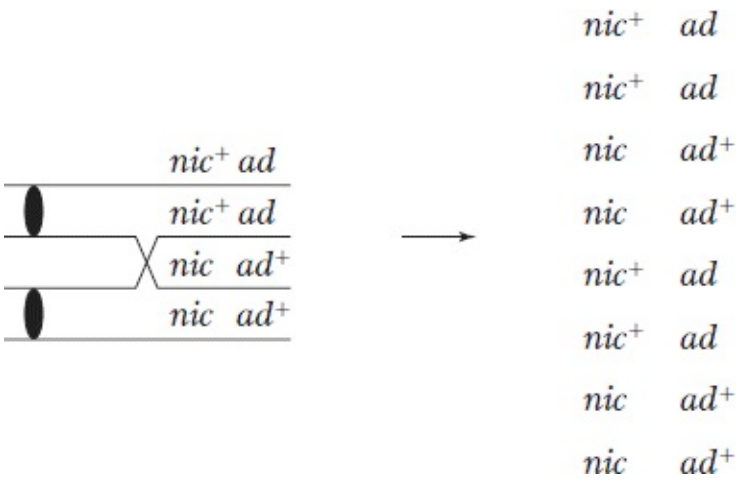


Agora precisamos posicionar os dois em conjunto e decidir entre as alternativas a seguir, todas as quais são compatíveis com as distâncias do *locus* até o centrômero precedentes:



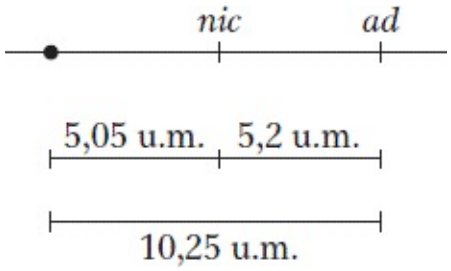
Aqui, uma combinação do senso comum e da análise simples nos informa qual alternativa está correta. Primeiramente, uma inspeção dos ascos revela que o tipo único mais comum é aquele rotulado 1, que contém mais de 80% de todos os ascos. Esse tipo contém apenas os genótipos $nic^+ \cdot ad$ e $nic \cdot ad^+$ e eles são genótipos *parentais*. Assim, sabemos que a recombinação é consideravelmente baixa e os *loci* certamente estão ligados. Isso descarta a alternativa **a**.

Agora considere a alternativa **c**. Se essa alternativa estivesse correta, um *crossover* entre o centrômero e o *locus nic* geraria não apenas um padrão de M_{II} em relação àquele *locus*, mas também um padrão de M_{II} em relação ao *locus ad*, tendo em vista que ele está mais distante do centrômero do que o *nic*. O padrão de asco produzido por um *crossover* entre *nic* e o centrômero na alternativa **c** deve ser



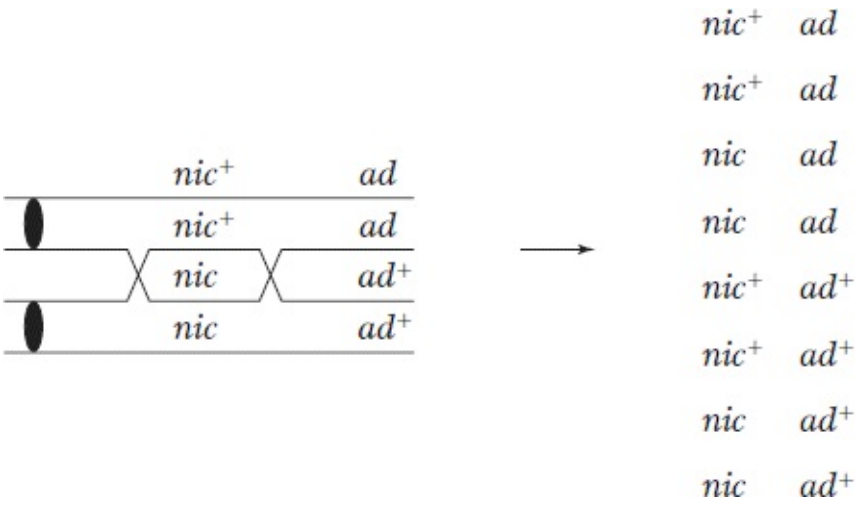
Relembre que o *locus nic* demonstra padrão de M_{II} nos ascos dos tipos 4, 5, 6 e 7 (um total de 101 ascos); deles, o tipo 5 é justamente aquele a respeito do qual estamos falando e contém 90 ascos. Portanto, a alternativa **c** aparenta ser a correta, tendo em vista que o asco do tipo 5 compreende aproximadamente 90% dos ascos de M_{II} em relação ao *locus nic*. Essa relação não seria mantida se a alternativa **b** estivesse correta, tendo em vista que *crossovers* em qualquer lado do centrômero gerariam padrões de M_{II} em relação os *loci nic* e *ad* de modo independente.

A distância de mapa de *nic* até *ad* é simplesmente 9,30 — 5,05 = 4,25 u.m.? Próximo disto, mas não tanto. O melhor modo de calcular as distâncias de mapa entre os *loci* sempre é por meio da medição da frequência de recombinantes. Podemos passar pelos ascos e contar todos os ascósporos recombinantes, mas com a utilização da fórmula FR = ½ T + DNP é mais simples. Os ascos T são das classes 3, 4 e 7, e os ascos DNP são das classes 2 e 6. Portanto, FR = [1/2 (100) + 2]/1.000 = 5,2%, ou 5,2 u.m., e um mapa melhor é:



O motivo da subestimativa da distância de *ad* até o centrômero calculada a partir

da frequência de M_{II} é a ocorrência de *crossovers* duplos, que podem produzir um padrão de M_I para *ad*, como no asco do tipo 4:



PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 4.3, existem quaisquer produtos meióticos que não passaram por um *crossover* na meiose ilustrada? Caso afirmativo, de quais cores eles seriam na convenção de cores utilizada?

2. Na Figura 4.5, por que o diagrama não demonstra meioses nas quais ocorrem dois *crossovers* entre as mesmas duas cromátides (tais como as duas internas)?

3. Na Figura 4.7, alguns produtos meióticos são rotulados parentais. A qual genitor está sendo feita referência nessa terminologia?

4. Na Figura 4.8, por que apenas o *locus A* está demonstrado em uma posição constante?

5. Na Figura 4.9, qual é a frequência média de *crossovers* por meiose na região *A*-*B*? Na região *B*-*C*?

6. Na Figura 4.10, é verdadeiro dizer que, a partir de um dado cruzamento, o produto $v\ cv^+$ pode apresentar duas origens diferentes?

7. Na Figura 4.13, na fileira inferior, quatro cores estão rotuladas SCO. Por

que elas não são todas do mesmo tamanho (frequência)?

8. Com a utilização das convenções da Figura 4.14, desenhe as classes dos genitores e das progênies a partir de um cruzamento:

P M'''/*p* M' × *p* M'/*p* M''''

9. Na Figura 4.16, desenhe os arranjos de alelos em uma óctade a partir de uma meiose semelhante na qual o produto superior da primeira divisão segregou de modo invertido na segunda divisão.

10. Na Figura 4.18, qual seria a FR entre *A/a* e *B/b* em um cruzamento no qual puramente ao acaso todas as meioses apresentassem *crossovers* duplos de quatro filamentos naquela região?

11.

a. Na Figura 4.20, GC = *A* e AT = *a*. Desenhe a óctade fúngica que resultaria da estrutura final (5).

b. (Desafiador) Insira alguns marcadores flanqueadores proximamente ligados no diagrama, digamos *P/p* à esquerda e *Q/q* à direita (presuma arranjos cis ou trans). Presuma que nenhum desses *loci* apresente segregação não mendeliana. Em seguida, desenhe a óctade final com base na estrutura na parte 5.

PROBLEMAS BÁSICOS

12. Uma planta de genótipo:

$$\frac{A \quad B}{a \quad b}$$

é submetida ao cruzamento-teste com:

$$\frac{a \quad b}{a \quad b}$$

Se os dois *loci* estiverem a uma distância de 10 u.m., qual proporção da progênie será *AB/ab*?

13. O *locus A* e o *locus D* estão ligados de modo tão próximo que nenhuma recombinação chega a ser observada entre eles. Se *Ad/Ad* for cruzado com *aD/aD* e a F_1 for entrecruzada, quais fenótipos serão observados na F_2 e em quais proporções?

14. Os *loci R* e *S* estão a uma distância de 35 u.m. Se uma planta de genótipo:

$$\frac{R \quad S}{r \quad s}$$

for autofecundada, quais fenótipos da progênie serão observados e em quais proporções?

15. É realizado o cruzamento $E/E \cdot F/F \times e/e \cdot f/f$ e em seguida a F_1 é retrocruzada com o genitor recessivo. Os genótipos da progênie são inferidos a partir dos fenótipos. Os genótipos da progênie, escritos como as contribuições gaméticas do genitor heterozigoto, estão nas proporções a seguir:

$E \cdot F$	$\frac{2}{6}$
$E \cdot f$	$\frac{1}{6}$
$e \cdot F$	$\frac{1}{6}$
$e \cdot f$	$\frac{2}{6}$

Explique esses resultados.

16. Uma linhagem de *Neurospora* com o genótipo $H \cdot I$ é cruzada com uma linhagem com o genótipo $h \cdot i$. Metade da progênie é $H \cdot I$ e a outra metade é $h \cdot i$. Explique como esse desfecho é possível.

17. Uma fêmea com genótipo $A/a \cdot B/b$ é cruzada com um macho duplo recessivo ($a/a \cdot b/b$). Sua progênie inclui 442 $A/a \cdot B/b$, 458 $a/a \cdot b/b$, 46 $A/a \cdot b/b$ e 54 $a/a \cdot B/b$. Explique esses resultados.

18. Se $A/A \cdot B/B$ for cruzado com $a/a \cdot b/b$ e a F_1 for submetida ao cruzamento-teste, qual porcentagem da progênie do cruzamento-teste será $a/a \cdot b/b$ se os dois genes forem **(a)** não ligados; **(b)** completamente ligados

(absolutamente nenhum *crossing over*); **(c)** distantes 10 u.m.; **(d)** distantes 24 u.m.?

19. Em um organismo haploide, os *loci C* e *D* estão a uma distância de 8 u.m. A partir de um cruzamento *C d* × *c D*, forneça a proporção de cada uma das classes de progênie a seguir: **(a)** *C D*; **(b)** *c d*; **(c)** *C d*; **(d)** todos os recombinantes combinados.

20. Uma mosca-das-frutas de genótipo *B R/b r* é submetida ao cruzamento-teste com *b r/b r*. Em 84% das meioses, não existem quiasmas entre os genes ligados; em 16% das meioses, existe um quiasma entre os genes. Qual proporção da progênie será *B r/b r*?

21. Foi realizado um cruzamento-teste de três pontos no milho. Os resultados e uma análise de recombinação estão demonstrados na ilustração a seguir, que é típica de cruzamentos-teste de três pontos (*p* = Folhas roxas, + = Verdes; *v* = Mudas resistentes a vírus, + = Sensíveis; *b* = Diafragma marrom na semente, + = Plano). Estude a ilustração e responda às partes *a* a *c*.

P $+/+ \cdot +/+ \cdot +/+ \times p/p \cdot v/v \cdot b/b$

Gametas $+ \cdot +$; $\cdot +$; $p \cdot v \cdot b$

a. Determine quais genes estão ligados.

b. Desenhe um mapa que demonstre as distâncias em unidades de mapa.

c. Calcule a interferência, se apropriado.

Como solucionar o problema 21

1. Esboce desenhos das plantas do milho P, da F_1 e testadoras, e utilize setas para demonstrar exatamente como você realizaria esse experimento. Demonstre onde as sementes são obtidas.
2. Por que todos os + apresentam aparência semelhante, até mesmo em relação a genes diferentes? Por que isso não causa confusão?

3. Como um fenótipo pode ser roxo e marrom, por exemplo, ao mesmo tempo?
4. É significativo que os genes sejam escritos na ordem *p-v-b* no problema?
5. Qual é um testador e por que ele é utilizado nessa análise?
6. O que a coluna marcada “Fenótipos da progênie” representa? Na classe 1, por exemplo, declare exatamente o que “verdes, sensíveis, plano” significa.
7. O que a linha marcada “Gametas” representa e como ela é diferente da coluna marcada “Gametas da F_1”? De que modo a comparação desses dois tipos de gametas é relevante para a recombinação?
8. Qual meiose é o principal foco do estudo? Rotule-a no seu desenho.
9. Por que os gametas do testador não estão demonstrados?
10. Por que existem apenas oito classes fenotípicas? Alguma classe está faltando?
11. Quais classes (e em quais proporções) seriam esperadas se todos os genes estivessem em cromossomos separados?
12. A que correspondem os quatro pares de tamanhos de classe (muito grande, duas intermediárias, muito pequenas)?
13. O que você pode dizer a respeito da ordem dos genes simplesmente ao inspecionar as classes fenotípicas e suas frequências?

Classe	Fenótipos da progênie	Gametas da F_1	Números	*Recombinante em rel* *a*	
				p-b	*p-v*
1	verdes, sensíveis, plano	+ · + · +	3.210		
2	roxas, resistentes,	*p* · *v* · *b*	3.222		

	marrom					
3	verdes, resistentes, plano	+ · *v* · +		1.024		R
4	roxas, sensíveis, marrom	*p* · + · *b*		1.044		R
5	roxas, resistentes, plano	*p* · *v* · +		690	R	
6	verdes, sensíveis, marrom	+ · + · *b*		678	R	
7	verdes, resistentes, marrom	+ · *v* · *b*		72	R	R
8	roxas, sensíveis, plano	*p* · + · +		60	R	R
			Total	10.000	1.500	2.200

14. Qual será a distribuição esperada das classes fenotípicas se apenas dois genes estiverem ligados?
15. A que a palavra “ponto” faz referência em um cruzamento-teste de três

pontos? A utilização dessa palavra implica ligação? Como seria um cruzamento-teste de quatro pontos?

16. Qual é a definição de *recombinante* e como ela é aplicada aqui?
17. O que as colunas “Recombinante em relação a” significam?
18. Por que existem apenas três colunas “Recombinante em relação a”?
19. O que os R significam e como eles são determinados?
20. O que os totais das colunas significam? Como eles são utilizados?
21. Qual é o teste diagnóstico para ligação?
22. O que é uma unidade de mapa? É o mesmo que um centimorgan?
23. Em um cruzamento-teste de três pontos como esse, por que a F_1 e o testador não são considerados como sendo parentais no cálculo da recombinação? (Eles *são* genitores em um sentido.)
24. Qual é a fórmula em relação à interferência? Como as frequências “esperadas” são calculadas na fórmula do coeficiente de coincidência?
25. Por que a parte *c* do problema diz “se apropriado”?
26. É muito trabalhoso obter um referido tamanho grande de progênie no milho? Qual dos três genes seria o mais trabalhoso para ser classificado? Aproximadamente quantas progênies são representadas por uma espiga?

22. Você possui uma linhagem de *Drosophila* que é homozigota em relação aos alelos recessivos *a, b* e *c*, ligados naquela ordem. Você cruza fêmeas dessa linhagem com machos homozigotos em relação aos alelos do tipo selvagem correspondentes. Em seguida você cruza os machos heterozigotos da F_1 com suas irmãs heterozigotas. Você obtém os fenótipos da F_2 a seguir (nos quais as letras indicam fenótipos recessivos e sinais de mais indicam fenótipos do tipo selvagem): 1.364 + + +, 365 *a b c*, 87 *a b* +, 84 + + *c*, 47 *a* + +, 44 + *b c*, 5 *a* + *c* e 4 + *b* +.

a. Qual é a frequência de recombinantes entre *a* e *b*? Entre *b* e *c*? (Lembre-se, não existe *crossover* em machos de *Drosophila*.)

b. Qual é o coeficiente de coincidência?

23. R. A. Emerson cruzou duas linhagens puras diferentes de milho e obteve uma F_1 fenotipicamente do tipo selvagem que era heterozigota em relação aos três alelos que determinam fenótipos recessivos: *an* determina antera; *br*, braquítico; e *f*, fino. Ele realizou o cruzamento-teste da F_1 com um testador que era homozigoto recessivo em relação aos três genes e obteve esses fenótipos de progênie: 355 anteras; 339 braquíticos e finos; 88 completamente do tipo selvagem; 55 anteras, braquíticos e finos; 21 finos; 17 anteras e braquíticos; 2 braquíticos; 2 anteras e finos.

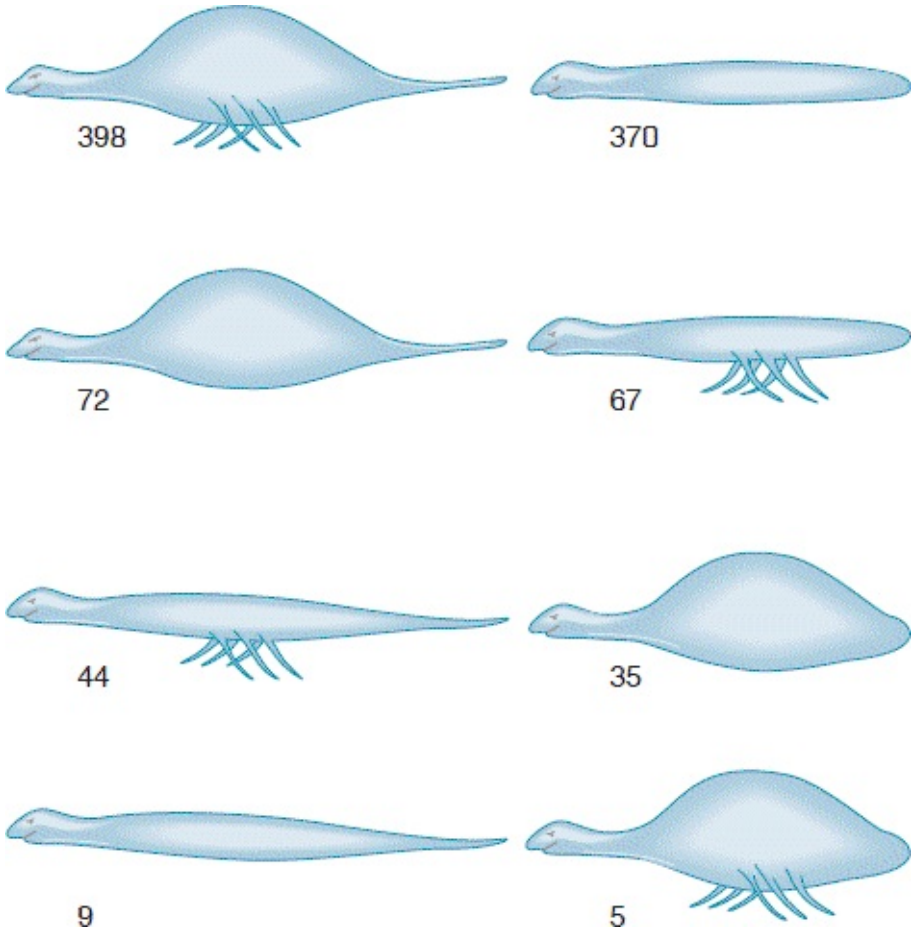
a. Quais eram os genótipos das linhagens parentais?

b. Desenhe um mapa de ligação em relação aos três genes (inclua as distâncias de mapa).

c. Calcule o valor da interferência.

24. O cromossomo 3 do milho carrega três *loci* (*b* para intensificador da cor da planta, *v* para virescente e *lg* para sem língula). Um cruzamento-teste de recessivos triplos com plantas da F_1 heterozigotas em relação aos três genes produz uma progênie que apresenta os genótipos a seguir: 305 + *v lg*, 275 *b* + +, 128 *b* + *lg*, 112 + *v* +, 74 + + *lg*, 66 *b v* +, 22 + + + e 18 *b v lg*. Forneça a sequência dos genes no cromossomo, as distâncias de mapa entre os genes e o coeficiente de coincidência.

25. *Groodies* são organismos haploides úteis (porém fictícios) que são puras ferramentas genéticas. Um *groody* do tipo selvagem apresenta um corpo gordo, uma cauda longa e flagelos. Sabidamente as linhagens mutantes apresentam corpos magros, não têm cauda e não apresentam flagelos. Os *groodies* podem ser cruzados entre si (embora sejam tão tímidos que não sabemos como) e produzem recombinantes. Um *groody* do tipo selvagem cruza com um *groody* magro e com ausência de cauda e de flagelos. Os 1.000 bebês *groodies* produzidos são classificados conforme demonstrado na ilustração aqui. Atribua os genótipos e mapeie os três genes. (O Problema 25 é de Burton S. Guttman.)



26. Em *Drosophila*, o alelo dp^+ determina asas longas e *dp* determina asas curtas ("atarracadas"). Em um *locus* separado, e^+ determina corpo cinza e *e* determina corpo cor de ébano. Ambos os *loci* são autossômicos. Foram realizados os cruzamentos a seguir, iniciando com genitores puros:

P ♀ longas e ébano × ♂ curtas e cinza

↓

F_1 ♀ longas e cinza × ♂ curtas e ébano (pura)

↓

F_2	longas e ébano	54
	longas e cinza	47
	curtas e cinza	52
	curtas e ébano	47
		200

Utilize o teste do χ^2 para determinar se esses *loci* estão ligados. Ao realizar isso, indique **(a)** a hipótese, **(b)** o cálculo do χ^2, **(c)** o valor *p*, **(d)** o que o valor *p* significa, **(e)** a sua conclusão, **(f)** a constituição cromossômica inferida dos genitores, da F_1, do testador e da progênie.

27. A mãe de uma família com 10 filhos apresenta tipo sanguíneo Rh^+. Ela também apresenta uma condição muito rara (eliptocitose, fenótipo E) que faz com que os eritrócitos tenham um formato oval em vez de redondo, mas que não produz efeitos clínicos adversos. O pai é Rh^- (não possui o antígeno Rh^+) e apresenta eritrócitos normais (fenótipo e). Os filhos são 1 Rh^+ e, 4 Rh^+ E, e 5 Rh^- e. Estão disponíveis informações sobre os genitores da mãe, que são Rh^+ E e Rh^- e. Um dos 10 filhos (que é Rh^+ E) se casa com uma pessoa que é Rh^+ e, e eles têm um filho Rh^+ E.

a. Desenhe o heredograma dessa família.

b. O heredograma está de acordo com a hipótese de que o alelo Rh^+ é dominante e o Rh^- é recessivo?

c. Qual é o mecanismo de transmissão da eliptocitose?

d. Os genes que controlam os fenótipos E e Rh podem estar no mesmo cromossomo? Caso afirmativo, estime a distância de mapa entre eles e comente o seu resultado.

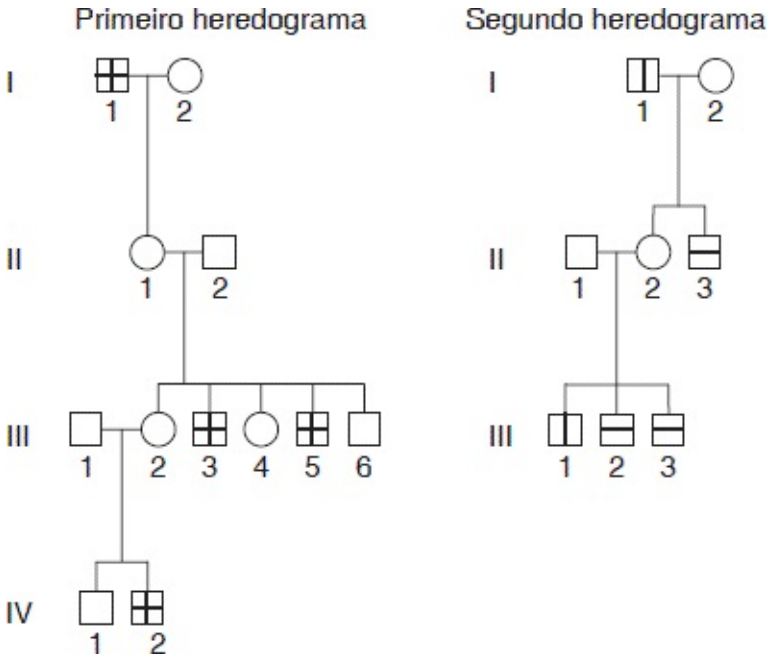
28. A partir de diversos cruzamentos do tipo geral $A/A \cdot B/B \times a/a \cdot b/b$, os indivíduos da F_1 do tipo $A/a \cdot B/b$ foram submetidos ao cruzamento-teste com $a/a \cdot b/b$. Os resultados são como segue:

Cruzamentoteste da F_1 do cruzamento	*Progênie do cruzamento-teste*			
	$A/a \cdot B/b$	$a/a \cdot b/b$	$A/a \cdot b/b$	$a/a \cdot B/b$
1	310	315	287	288
2	36	38	23	23

3	360	380	230	230
4	74	72	50	44

Em relação a cada grupo de progênie, utilize o teste do χ^2 para decidir se existe evidência de ligação.

29. Nos dois heredogramas diagramados aqui, uma barra vertical em um símbolo faz referência à deficiência de esteroide sulfatase e uma barra horizontal faz referência à deficiência de ornitina transcarbamilase.



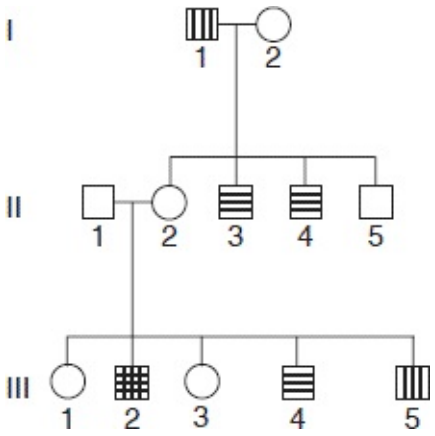
a. Existe evidência nesses heredogramas de que os genes que determinam as deficiências estão ligados?

b. Se os genes estão ligados, existem evidências no heredograma de *crossover* entre eles?

c. Atribua os genótipos desses indivíduos tanto quanto possível.

30. No heredograma a seguir, as linhas verticais fazem referência ao daltonismo protano e as horizontais, ao daltonismo deutano. Essas são condições separadas, que causam percepções errôneas diferentes das

cores; cada uma é determinada por um gene separado.



a. O heredograma demonstra evidências de que os genes estão ligados?

b. Se houver ligação, como o heredograma demonstra evidências de *crossing over*?

Explique as suas respostas em relação às partes *a* e *b* com o auxílio do diagrama.

c. Você consegue calcular um valor para a recombinação entre esses genes? Essa recombinação ocorre por meio de distribuição independente ou por meio de *crossover*?

31. No milho, foi obtido um heterozigoto triplo que carrega os alelos mutantes *s* (encolhido), *w* (aleurona branca) e *y* (endosperma ceroso), todos pareados com seus alelos do tipo selvagem normais. Esse heterozigoto triplo foi submetido ao cruzamento-teste e a progênie continha 116 encolhidos e brancos; 4 totalmente do tipo selvagem; 2.538 encolhidos; 601 encolhidos e cerosos; 626 brancos; 2.708 brancos e cerosos; 2 encolhidos, brancos e cerosos; e 113 cerosos.

a. Determine se quaisquer desses três *loci* estão ligados e, caso positivo, demonstre as distâncias de mapa.

b. Demonstre o arranjo de alelos nos cromossomos do heterozigoto triplo utilizado no cruzamento-teste.

c. Calcule a interferência, se apropriado.

32.

a. É realizado um cruzamento de camundongos *A/a · B/b* × *a/a · b/b* e na progênie existe:

25% *A/a · B/b*, 25% *a/a · b/b*,
25% *A/a · b/b*, 25% *a/a · B/b*

Explique essas proporções com o auxílio de diagramas de meiose simplificados.

b. É realizado um cruzamento de camundongos *C/c · D/d* × *c/c · d/d* e na progênie existe:

45% *C/c · d/d*, 45% *c/c · D/d*,
5% *c/c · d/d*, 5% *C/c ·D/d*

Explique essas proporções com o auxílio de diagramas de meiose simplificados.

33. Na pequena planta-modelo *Arabidopsis*, o alelo recessivo *hyg* confere resistência às sementes ao fármaco higromicina e *her*, um alelo recessivo de um gene diferente, confere resistência a herbicidas. Uma planta que era homozigota *hyg/hyg · her/her* foi cruzada com o tipo selvagem e a F_1 foi autofecundada. As sementes resultantes da F_1 autofecundada foram colocadas em placas de Petri contendo higromicina e herbicida.

a. Se os dois genes não forem ligados, espera-se o crescimento de qual porcentagem de sementes?

b. De fato, 13% das sementes cresceram. Essa porcentagem ampara a hipótese de não ligação? Explique. Caso negativo, calcule o número de unidades de mapa entre os *loci*.

c. Sob a sua hipótese, se a F_1 for submetida ao cruzamento-teste, qual proporção das sementes crescerá no meio que contém higromicina e herbicida?

34. Em um organismo diploide de genótipo *A/a; B/b; D/d*, os pares de alelos estão todos em diferentes pares de cromossomos. O diagrama ao lado demonstra anáfases (estágios de “separação”) em células individuais.

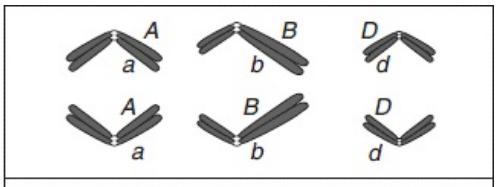
Declare se cada desenho representa mitose, meiose I ou meiose II ou se isso é impossível em relação a esse genótipo em particular.

35. É realizado o *cruzamento de Neurospora al*-2^{+} × *al*-2. Uma análise de tétrade linear revela que a frequência de segregação da segunda divisão é de 8%.

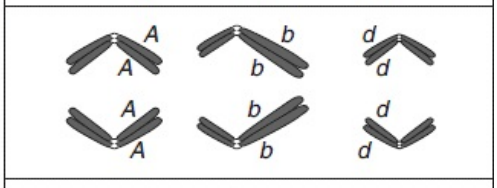
a. Desenhe dois exemplos de padrões de segregação de segunda divisão nesse cruzamento.

b. O que pode ser calculado por meio da utilização do valor de 8%?

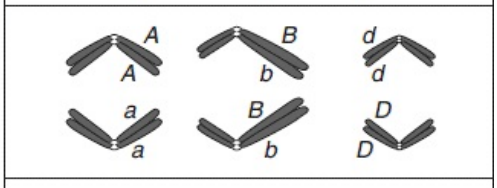
a.


A
a
B
b
D
d
A
a
B
b
D
d

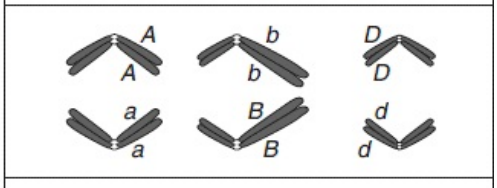
b.


A
A
b
b
d
d
A
A
b
b
d
d

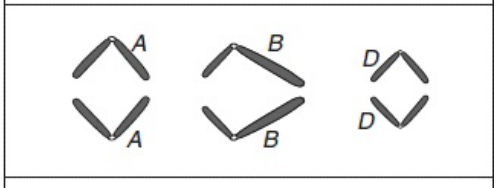
c.


A
A
B
b
d
d
a
a
B
b
D
D

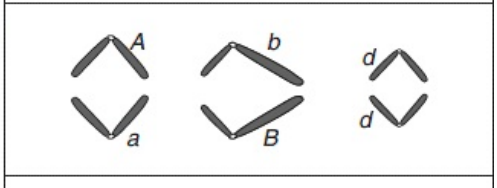
d.


A
A
b
b
D
D
a
a
B
B
d
d

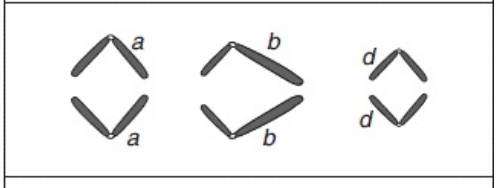
e.


A
A
B
B
D
D

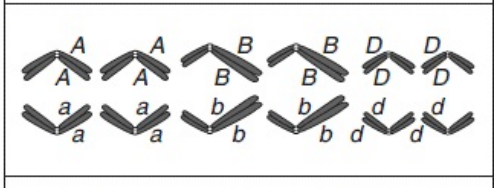
f.


A
a
b
B
d
d

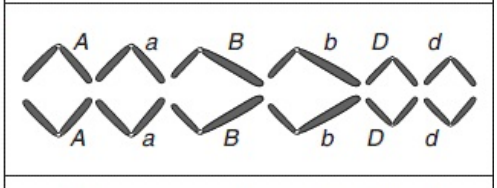
g.


a
a
b
b
d
d

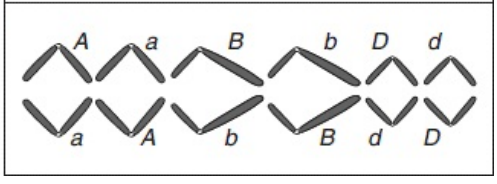
h.


A
A
A
A
B
B
B
B
D
D
D
D
a
a
a
a
b
b
b
b
d
d
d
d

i.


A
a
B
b
D
d
A
a
B
b
D
d

j.


A
a
B
b
D
d
a
A
b
B
d
D

36. A partir do cruzamento de fungos *arg-6* · *al*-2 × *arg-6*$^{+}$ · *al*-2^{+}, quais serão os genótipos de esporos em tétrades não ordenadas que são: **(a)** ditipos parentais? **(b)** tetratipos? **(c)** ditipos não parentais?

37. Em relação a uma determinada região cromossômica, o número médio de *crossovers* na meiose é calculado como sendo de dois por meiose. Naquela região, qual proporção de meioses se espera que apresente: **(a)** nenhum *crossover*? **(b)** um *crossover*? **(c)** dois *crossovers*?

38. Foi realizado um cruzamento de *Neurospora* entre uma linhagem que carregava o alelo de tipo reprodutivo *A* e o alelo mutante *arg-1* e outra linhagem que carregava o alelo de tipo reprodutivo *a* e o alelo do tipo selvagem em relação a *arg-1* (+). Foram isoladas quatrocentas óctades lineares e elas se situaram em sete classes fornecidas na tabela a seguir. (Para simplicidade, elas são demonstradas como tétrades.)

a. Deduza o arranjo de ligação do *locus* do tipo reprodutivo e do *locus arg-1*. Inclua o centrômero ou os centrômeros em qualquer mapa que você desenhe. Rotule *todos* os intervalos em unidades de mapa.

b. Diagrame as divisões meióticas que levaram à classe 6. Rotule claramente.

1	2	3	4	5	6	7
A · *arg*	*A* · +	*A* · *arg*	*A* · *arg*	*A* · *arg*	*A* · +	*A* · +
A · *arg*	*A* · +	*A* · +	*a* · *arg*	*a* · +	*a* · *arg*	*a* · *arg*
A · +	*a* · *arg*	*a* · *arg*	*A* · +	*A* · *arg*	*A* · +	*A* · *arg*
			a ·			

A · +	*a* · *arg*	*a* · +	+	*a* · +	*a* · *arg*	*a* · +
127	125	100	36	2	4	6

Como solucionar o problema 38

1. Os fungos em geral são haploides ou diploides?
2. Quantos ascósporos estão no asco de *Neurospora*? A sua resposta corresponde ao número apresentado neste problema? Explique qualquer discrepância.
3. Qual é o tipo reprodutivo em fungos? Como você acha que ele é determinado experimentalmente?
4. Os símbolos *A* e *a* têm alguma relação com dominância e recessividade?
5. O que o símbolo *arg-1* significa? Como você testaria em relação a esse genótipo?
6. Como o símbolo *arg-1* está relacionado ao símbolo +?
7. O que a expressão *do tipo selvagem* significa?
8. O que a palavra *mutante* significa?
9. A função biológica dos alelos demonstrada tem alguma relação com a solução deste problema?
10. O que a expressão *análise de óctade linear* significa?
11. Em geral, o que mais pode ser aprendido a partir da análise de tétrade linear que não pode ser aprendido a partir da análise de tétrade não ordenada?
12. Como é realizado um cruzamento em um fungo como *Neurospora*? Explique como isolar ascos e ascósporos individuais. Como o termo *tétrade* se relaciona com os termos *asco* e *óctade*?
13. Onde ocorre a meiose no ciclo de vida de *Neurospora*? (Demonstre-a no

diagrama do ciclo de vida.)

14. Qual a relação do Problema 38 com a meiose?
15. Você consegue escrever os genótipos das duas linhagens parentais?
16. Por que somente quatro genótipos estão demonstrados em cada classe?
17. Por que existem apenas sete classes? Quantos outros modos você aprendeu para a classificação das tétrades em geral? Qual dessas classificações pode ser aplicada a ambas as tétrades lineares e não ordenadas? Você consegue aplicar essas classificações às tétrades neste problema? (Classifique cada classe de tantos modos quanto possível.) Você consegue pensar sobre mais possibilidades neste cruzamento? Caso afirmativo, por que elas não estão demonstradas?
18. Você acredita que existam diversas diferentes ordens de esporos dentro de cada classe? Por que essas diferentes ordens de esporos não alteram a classe?
19. Por que a classe a seguir não está listada?

$$a \cdot + \qquad A \cdot arg$$
$$a \cdot + \qquad A \cdot arg$$

20. O que a expressão *arranjo de ligação* significa?
21. O que é um *intervalo* genético?
22. Por que o problema declara “centrômero ou centrômeros”, não apenas “centrômero”? Qual é o método geral para o mapeamento de centrômeros na análise de tétrades?
23. Qual é a frequência total de ascósporos $A \cdot +$? (Você calculou essa frequência por meio da utilização de uma fórmula ou por meio da inspeção? Esse é um genótipo recombinante? Caso afirmativo, ele é o único genótipo recombinante?)
24. As primeiras duas classes são as mais comuns e são aproximadamente iguais em frequência. O que essa informação lhe diz? Qual é o seu conteúdo

de genótipos parentais e recombinantes?

39. Um geneticista estuda 11 diferentes pares de *loci* de *Neurospora* ao realizar cruzamentos do tipo $a \cdot b \times a^+ \cdot b^+$ e em seguida ao analisar 100 ascos lineares de cada cruzamento. Para a conveniência da confecção de uma tabela, o geneticista organiza os dados como se todos os 11 pares de genes apresentassem a mesma designação — *a* e *b* — conforme demonstrado a seguir. Em relação a cada cruzamento, mapeie os *loci* em relação uns aos outros e aos centrômeros.

	Número de ascos do tipo						
Cruzamento	$a \cdot b$ / $a \cdot b$	$a \cdot b^+$ / $a \cdot b^+$	$a \cdot b$ / $a \cdot b^+$	$a \cdot b$ / $a^+ \cdot b$	$a \cdot b$ / $a^+ \cdot b^+$	$a \cdot b^+$ / $a^+ \cdot b$	$a \cdot b^+$ / $a^+ \cdot b$
	$a^+ \cdot b^+$ / $a^+ \cdot b^+$	$a^+ \cdot b$ / $a^+ \cdot b$	$a^+ \cdot b^+$ / $a^+ \cdot b$	$a^+ \cdot b^+$ / $a \cdot b^+$	$a^+ \cdot b^+$ / $a^+ \cdot b^+$	$a^+ \cdot b$ / $a^+ \cdot b$	$a^+ \cdot b^+$ / $a \cdot b$
1	34	34	32	0	0	0	0
2	84	1	15	0	0	0	0
3	55	3	40	0	2	0	0
4	71	1	18	1	8	0	1

5	9	6	24	22	8	10	20
6	31	0	1	3	61	0	4
7	95	0	3	2	0	0	0
8	6	7	20	22	12	11	22
9	69	0	10	18	0	1	2
10	16	14	2	60	1	2	5
11	51	49	0	0	0	0	0

40. Três cruzamentos diferentes em *Neurospora* são analisados com base em tétrades não ordenadas. Cada cruzamento combina um par diferente de genes ligados. Os resultados estão demonstrados na tabela a seguir:

Cruzamento	Genitores (%)	Ditipos parentais (%)	Tetratipos (%)	Ditipos não parentais (%)
1	$a \cdot b^{+} \times a^{+} \cdot b$	51	45	4
2	$c \cdot d^{+} \times c^{+} \cdot d$	64	34	2
3	$e \cdot f^{+} \times e^{+} \cdot f$	45	50	5

Em relação a cada cruzamento, calcule:

a. A frequência de recombinantes (FR).

b. A distância de mapa não corrigida, com base na FR.

c. A distância de mapa corrigida, com base nas frequências de tétrades.

d. A distância de mapa corrigida, com base na função de mapeamento.

41. No cromossomo 4 de *Neurospora,* o gene *leu3* está logo à esquerda do centrômero e sempre segrega na primeira divisão, enquanto o gene *cys2* está à direita do centrômero e demonstra uma frequência de segregação na segunda divisão de 16%. Em um cruzamento entre uma linhagem *leu3* e uma cepa *cys2*, calcule as frequências previstas das sete classes a seguir de tétrades lineares nas quais *l = leu 3* e *c = cys2*. (Ignore os *crossovers* duplos e outros *crossovers* múltiplos.)

(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
l c	*l* +	*l* +	+ *c*	+ +	+ *c*	+ *c*
l c	+ *c*	+ +	+ +	+ +	+ *c*	+ +
+ +	+ *c*	+ *c*	*l* +	*l c*	*l* +	*l c*
+ +	+ *c*	+ *c*	*l* +	*l c*	*l* +	*l c*

42. Um cultivador de arroz obteve um heterozigoto triplo que carrega os três alelos recessivos em relação a flores albinas (*al*), arestas marrons (*b*) e folhas felpudas (*fu*), todos pareados com seus alelos do tipo selvagem normais. Esse heterozigoto triplo foi submetido ao cruzamento-teste. Os fenótipos da progênie foram:

170	do tipo selvagem	710	albinos
150	albinos, marrons e felpudos	698	marrons e felpudos
5	marrons	42	felpudos
3	albinos e felpudos	38	albinos e marrons

a. Alguns dos genes estão ligados? Caso afirmativo, desenhe um mapa rotulado com as distâncias de mapa. (Não se preocupe com uma correção em relação a *crossovers* múltiplos.)

b. O heterozigoto triplo foi originalmente produzido por meio do cruzamento de duas linhagens puras. Quais são os seus genótipos?

43. Em um fungo, um mutante prolina (*pro*) foi cruzado com um mutante histidina (*his*). Uma análise de uma tétrade não linear produziu os resultados a seguir:

+	+	+	+	+	*his*
+	+	+	*his*	+	*his*
pro	*his*	*pro*	+	*pro*	+
pro	*his*	*pro*	*his*	*pro*	+
6		82		112	

a. Os genes estão ligados ou não?

b. Desenhe um mapa (se ligados) ou dois mapas (se não ligados) demonstrando as distâncias de mapa com base na frequência de recombinantes direta, onde apropriado.

c. Se houver ligação, corrija as distâncias de mapa em relação a *crossovers* múltiplos (escolha apenas uma abordagem).

44. No fungo *Neurospora*, uma linhagem que é auxotrófica em relação à tiamina (alelo mutante *t*) foi cruzada com uma linhagema que é auxotrófica em relação à metionina (alelo mutante *m*). Os ascos lineares foram isolados e classificados nos grupos a seguir:

Par de esporos	Tipos de ascos					
1 e 2	*t* +	*t* +	*t* +	*t* +	*t m*	*t m*
3 e 4	*t* +	*t m*	+ *m*	+ +	*t m*	+ +
5 e 6	+ *m*	+ +	*t* +	*t m*	+ +	*t* +
7 e 8	+ *m*	+ *m*	+	+ *m*	+ +	+ *m*

	m					
Número	260	76	4	54	1	5

a. Determine as relações de ligação desses dois genes com o(s) seu(s) centrômero(s) e entre si. Especifique as distâncias em unidades de mapa.

b. Desenhe um diagrama para demonstrar a origem do tipo de asco com apenas um representante único (segundo a partir da direita).

45. Uma geneticista deseja obter uma planta de milho que apresente os três fenótipos dominantes: antocianina (A), cabelos longos (L) e planta anã (D). Em sua coleção de linhagens puras, as únicas linhagens que contêm esses alelos são *AA LL dd* e *aa ll DD*. Ela também possui a linhagem totalmente recessiva *aa ll dd*. Ela decide entrecruzar as duas primeiras e realizar o cruzamento-teste do híbrido resultante para obter na progênie uma planta do fenótipo desejado (que nesse caso deve ser *Aa Ll Dd*). Ela sabe que os três genes estão ligados na ordem escrita, que a distância entre os *loci A/a* e *L/l* é de 16 u.m. e que a distância entre os *loci L/l e D/d* é de 24 u.m.

a. Desenhe um diagrama dos cromossomos dos genitores, do híbrido e do testador.

b. Desenhe um diagrama do(s) *crossover*(s) necessário(s) para produzir o genótipo desejado.

c. Qual porcentagem da progênie do cruzamento-teste será do fenótipo que ela necessita?

d. Quais presunções você fez (se existentes)?

46. Na planta-modelo *Arabidopsis thaliana*, os alelos a seguir foram utilizados em um cruzamento:

T	=	Presença de tricomas	*t*	=	Ausência de tricomas
D	=	Plantas altas	*d*	=	Plantas anãs
W	=	Cutícula cerosa	*w*	=	Não cerosa
A	=	Presença do pigmento roxo	*a*	=	Ausência (branca)

antocianina

Os *loci T/t* e *D/d* estão ligados a uma distância de 26 u.m. no cromossomo 1, enquanto os *loci W/w* e *A/a* estão ligados a uma distância de 8 u.m. no cromossomo 2.

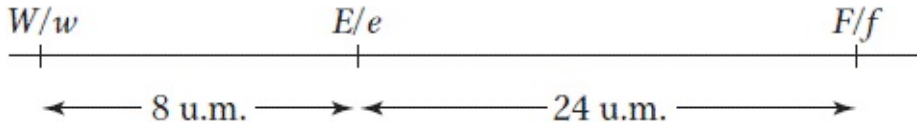
Uma planta não cerosa sem tricomas homozigota recessiva dupla pura é cruzada com outra planta branca anã homozigota recessiva dupla pura.

a. Qual será o aspecto da F_1?

b. Esboce os cromossomos 1 e 2 dos genitores e da F_1, demonstrando o arranjo dos alelos.

c. Se a F_1 for submetida ao cruzamento-teste, qual proporção da progênie apresentará todos os quatro fenótipos recessivos?

47. No milho, é realizado o cruzamento *WW ee FF* × *ww EE ff*. Os três *loci* estão ligados como segue:



Presuma a ausência de interferência.

a. Se a F_1 for submetida ao cruzamento-teste, qual proporção da progênie será *ww ee ff*?

b. Se a F_1 for autofecundada, qual proporção da progênie será *ww ee ff*?

48. Foi realizado o cruzamento de fungos + · + × *c* · *m* e foram coletadas tétrades *não lineares* (não ordenadas). Os resultados foram:

+ +	+ +	+ *m*
+ +	+ *m*	+ *m*
c m	*c* +	*c* +
c m	*c m*	*c* +

Total	112	82	6

a. A partir desses resultados, calcule uma frequência de recombinantes simples.

b. Compare a função de mapeamento de Haldane e a fórmula de Perkins em suas conversões do valor da FR em uma distância de mapa "corrigida".

c. Na derivação da fórmula de Perkins, apenas a possibilidade de meioses com zero, um e dois *crossovers* foi considerada. Esse limite poderia explicar qualquer discrepância nos seus valores calculados? Explique brevemente (nenhum cálculo é necessário).

49. Em camundongos, os alelos a seguir foram utilizados em um cruzamento:

W =	Valsador (deambulação	*w* =	Não valsador em valsa)
G =	Cor cinza normal	*g* =	Albino
B =	Cauda dobrada	*b* =	Cauda reta

Um camundongo cinza, de cauda dobrada e valsador é cruzado com um camundongo albino de cauda reta e não valsador e, ao longo de diversos anos, são obtidos os totais de progênie a seguir:

valsador	cinza	dobrada	18
valsador	albino	dobrada	21
não valsador	cinza	reta	19
não valsador	albino	reta	22
valsador	cinza	reta	4
valsador	albino	reta	5
não valsador	cinza	dobrada	5
não valsador	albino	dobrada	6
Total			100

a. Quais eram os genótipos dos dois camundongos parentais no cruzamento?

b. Desenhe os cromossomos dos genitores.

c. Se você deduziu ligação, declare o valor ou os valores de unidade de mapa e demonstre como eles foram obtidos.

50. Considere o cruzamento de *Neurospora* +; + × *f*; *p*.

Sabidamente o *locus* +/*f* está muito próximo do centrômero no cromossomo 7; na verdade, tão próximo que nunca ocorrem quaisquer segregações na segunda divisão. Também se sabe que o *locus* +/*p* está no cromossomo 5, a uma tal distância que normalmente ocorre uma média de 12% de segregações na segunda divisão. Com essas informações, qual será a proporção de óctades que são:

a. Ditipos parentais que demonstram padrões de M_I em relação a ambos os *loci*?

b. Ditipos não parentais que demonstram padrões de M_I em relação a ambos os *loci*?

c. Tetratipos que demonstram um padrão de M_I em relação a +/*f* e um padrão de M_{II} em relação a +/*p*?

d. Tetratipos que demonstram um padrão de M_{II} em relação a +/*f* e um padrão de M_I em relação a +/*p*?

51. Em um fungo haploide, os genes *al-2* e *arg-6* estão a uma distância de 30 u.m. no cromossomo 1 e os genes *lys-5* e *met-1* estão a uma distância de 20 u.m. no cromossomo 6. Em um cruzamento:

al-2 +; + *met-1* × + *arg-6*; *lys-5* +

qual proporção da progênie será prototrófica + +; + +?

52. Os alelos recessivos *k* (olhos com formato de rim em vez de redondos do tipo selvagem), *c* (cor dos olhos cardeal em vez do tipo selvagem vermelho) e *e* (corpo cor de ébano em vez do tipo selvagem cinza) identificam três genes no cromossomo 3 de *Drosophila*. Fêmeas com olhos em formato de rim e cor cardeal foram cruzadas com machos cor de ébano. A F_1 foi do tipo selvagem. Quando as fêmeas da F_1 foram submetidas ao cruzamento-teste com machos *kk cc ee*, foram obtidos os fenótipos da

progênie a seguir:

k	*c*	*e*	3
k	*c*	+	876
k	+	*e*	67
k	+	+	49
+	*c*	*e*	44
+	*c*	+	58
+	+	*e*	899
+	+	+	4
		Total	2.000

a. Determine a ordem dos genes e as distâncias de mapa entre eles.

b. Desenhe os cromossomos dos genitores e da F_1.

c. Calcule a interferência e diga o que você pensa a respeito da sua significância.

53. A partir de genitores com genótipos *A/A · B/B* e *a/a · b/b*, foi produzido um di-híbrido. Em um cruzamento-teste do di-híbrido, foi obtida a progênie de sete indivíduos a seguir:

A/a · B/b, *a/a · b/b*, *A/a · B/b*, *A/a · b/b*,
a/a · b/b, *A/a · B/b* e *a/a · B/b*

Esses resultados fornecem evidências convincentes de ligação?

PROBLEMAS DESAFIADORES

54. Utilize a função de mapeamento de Haldane para calcular a distância de mapa corrigida em casos nos quais a FR medida = 5%, 10%, 20%, 30% e 40%. Esboce um gráfico de FR em face da distância de mapa corrigida e utilize-o para responder a questão: Quando se deve utilizar a função de mapeamento?

Como solucionar o Problema 55

55. Um indivíduo heterozigoto em relação a quatro genes, *A/a · B/b · C/c · D/d*, é submetido ao cruzamento-teste com *a/a · b/b · c/c · d/d* e uma progênie de 1.000 foi classificada pela contribuição gamética do genitor heterozigoto como segue:

a · B · C · D	42
A · b · c · d	43
A · B · C · d	140
a · b · c · D	145
a · B · c · D	6
A · b · C · d	9
A · B · c · d	305
a · b · C · D	310

a. Quais genes estão ligados?

b. Se duas linhagens puras fossem cruzadas para produzir o indivíduo heterozigoto, quais teriam sido os seus genótipos?

c. Desenhe um mapa de ligação dos genes ligados, demonstrando a ordem e as distâncias em unidades de mapa.

d. Calcule um valor de interferência, se apropriado.

56. Um alelo autossômico *N* em seres humanos causa anormalidades nas unhas e patelas (rótulas), denominadas síndrome unha-patela. Considere casamentos nos quais um parceiro apresenta a síndrome unha-patela e o tipo sanguíneo A e o outro parceiro apresenta unhas e patelas normais e tipo sanguíneo 0. Esses casamentos produzem alguns filhos que apresentam ambas a síndrome unha-patela e o tipo sanguíneo A. Presuma que filhos não relacionados desse grupo fenotípico crescem, se casam entre si e têm filhos. São observados quatro fenótipos, nas porcentagens a seguir, nessa segunda geração:

Síndrome unha-patela e tipo sanguíneo A	66%
Unhas e patelas normais e tipo sanguíneo 0	16%
Unhas e patelas normais e tipo sanguíneo A	9%
Síndrome unha-patela e tipo sanguíneo 0	9%

Analise totalmente esses dados, explicando as frequências relativas dos quatro fenótipos. (Ver Capítulo 6 para a base genética desses tipos sanguíneos.)

57. Presuma que três pares de alelos sejam observados na *Drosophila*: x^+ e x, y^+ e y e z^+ e z. Conforme demonstrado pelos símbolos, cada alelo não do tipo selvagem é recessivo do seu alelo do tipo selvagem. Um cruzamento entre fêmeas heterozigotas nesses três *loci* e machos do tipo selvagem produz progênie que apresenta os genótipos a seguir: 1.010 fêmeas $x^+ \cdot y^+ \cdot z^+$, 430 machos $x \cdot y^+ \cdot z$, 441 machos $x^+ \cdot y \cdot z^+$, 39 machos $x \cdot y \cdot z$, 32 machos $x^+ \cdot y^+ \cdot z$, 30 machos $x^+ \cdot y^+ \cdot z^+$, 27 machos $x \cdot y \cdot z^+$, 1 macho $x^+ \cdot y \cdot z$ e 0 macho $x \cdot y^+ \cdot z^+$.

a. Em qual cromossomo de *Drosophila* estão esses genes?

b. Desenhe os cromossomos relevantes na genitora heterozigota, demonstrando o arranjo dos alelos.

c. Calcule as distâncias de mapa entre os genes e o coeficiente de coincidência.

58. Os cinco conjuntos de dados fornecidos na tabela a seguir representam os resultados de cruzamentos-teste com a utilização de genitores com os mesmos alelos, mas em diferentes combinações. Determine a ordem dos genes por meio da inspeção — ou seja, sem calcular os valores de recombinação. Os fenótipos recessivos estão simbolizados pelas letras minúsculas e os dominantes por sinais +.

Fenótipos observados em cruzamento-teste de 3 pontos	*Conjuntos de dados*				
	1	2	3	4	5
+ + +	317	1	30	40	305
+ + *c*	58	4	6	232	0
+ *b* +	10	31	339	84	28
+ *b c*	2	77	137	201	107
a + +	0	77	142	194	124
a + *c*	21	31	291	77	30
a b +	72	4	3	235	1
a b c	203	1	34	46	265

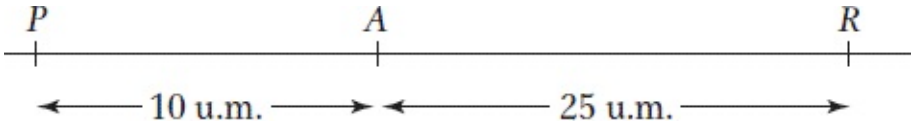
59. A partir dos dados de fenótipos fornecidos na tabela a seguir para cruzamentos-teste de 3 pontos em relação a (1) *a*, *b* e *c* e (2) *b*, *c* e *d*, determine a sequência dos quatro genes, *a*, *b*, *c* e *d* e as três distâncias de mapa entre eles. Os fenótipos recessivos estão simbolizados por letras minúsculas e os fenótipos dominantes por +.

1		2	
+ + +	669	*b c d*	8
a b +	139	*b* + +	441

a + +	3	*b* + *d*	90
+ + *c*	121	+ *c d*	376
+ *b c*	2	+ + +	14
a + *c*	2.280	+ + *d*	153
a b c	653	+ *c* +	65
+ *b* +	2.215	*b c* +	141

---

60. Vulcanos apresentam orelhas pontudas (determinadas pelo alelo *P*), ausência de adrenais (determinada por *A*) e coração do lado direito (determinado por *R*). Todos esses alelos são dominantes em relação aos alelos terráqueos normais: orelhas redondas (p), existência de adrenais (a) e coração do lado esquerdo (r). Os três *loci* são autossômicos e estão ligados conforme demonstrado nesse mapa de ligação:



O Sr. Spock, o primeiro oficial da espaçonave *Enterprise*, possui pai vulcano e mãe terráquea. Se o Sr. Spock se casar com uma terráquea e não houver interferência (genética), qual proporção de seus filhos apresentará:

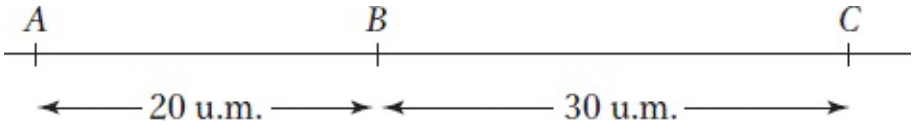
a. Fenótipos vulcanos em relação às três características?

b. Fenótipos terráqueos em relação às três características?

c. Orelhas e coração vulcanos, mas adrenais terráqueas?

d. Orelhas vulcanas, mas coração e adrenais terráqueos?

61. Em uma determinada planta diploide, os três *loci A, B* e *C* estão ligados como segue:



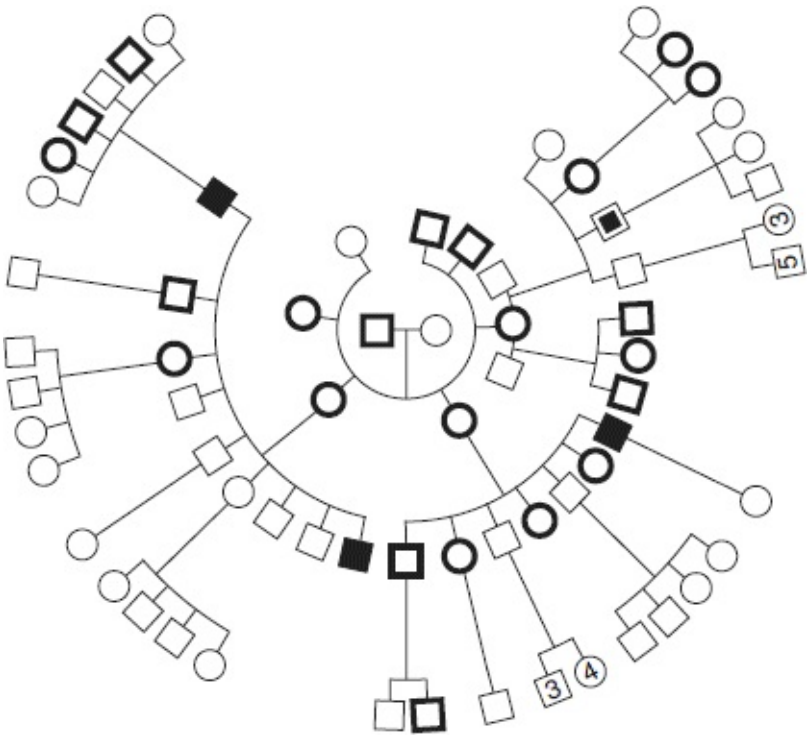
Uma planta está disponível para você (denomine-a planta genitora). Ela apresenta a constituição *A b c/a B C*.

a. Com a presunção de ausência de interferência, se a planta for autofecundada, qual proporção da progênie será do genótipo *a b c/a b c*?

b. Novamente, com a presunção de ausência de interferência, se a planta genitora for cruzada com a planta *a b c/a b c*, quais classes genotípicas serão observadas na progênie? Quais serão as suas frequências se houver uma progênie de 1.000?

c. Repita a parte *b*, dessa vez presumindo 20% de interferência entre as regiões.

62. O heredograma a seguir demonstra uma família com dois fenótipos anormais raros: esclerótica azul (associada à osteogênese imperfeita), representado por um símbolo com bordas pretas, e hemofilia, representada por um centro preto no símbolo. Os membros representados por símbolos completamente pretos apresentam ambos os distúrbios. Os números em alguns símbolos são os números de indivíduos com aqueles tipos.



a. Qual padrão de herança é demonstrado por cada condição nesse heredograma?

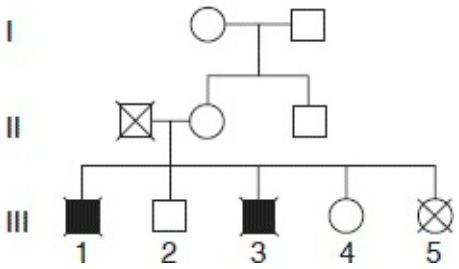
b. Forneça os genótipos de tantos familiares quanto possível.

c. Existe evidência de ligação?

d. Existe evidência de distribuição independente?

e. Algum dos familiares pode ser considerado como recombinante (ou seja, formado a partir de no mínimo um gameta recombinante)?

63. Ambos os genes humanos em relação ao daltonismo e à hemofilia estão no cromossomo X e demonstram uma frequência de recombinantes de aproximadamente 10%. A ligação de um gene patológico com um gene relativamente inofensivo pode ser utilizada para prognóstico genético. Aqui está demonstrada uma parte de um heredograma maior. Os símbolos escuros indicam que os indivíduos apresentam hemofilia e as cruzes indicam daltonismo. Quais informações podem ser fornecidas para as mulheres III-4 e III-5 a respeito da probabilidade de terem filhos com hemofilia?



(O Problema 63 é adaptado de J. F. Crow, *Genetics Notes: An Introduction to Genetics*. Burgess, 1983.)

64. Um geneticista que está mapeando os genes *A*, *B*, *C*, *D* e *E* realiza dois cruzamentos-teste de 3 pontos. O primeiro cruzamento de linhagens puras é:

A/A · B/B · C/C · D/D · E/E × *a/a · b/b · C/C · d/d · E/E*

O geneticista cruza a F_1 com um testador recessivo e classifica a progênie pela contribuição gamética da F_1:

A · B · C · D · E	316
a · b · C · d · E	314
A · B · C · d · E	31
a · b · C · D · E	39
A · b · C · d · E	130
a · B · C · D · E	140
A · b · C · D · E	17
a · B · C · d · E	13
	1.000

O segundo cruzamento de linhagens puras é *A/A · B/B · C/C · D/D · E/E* × *a/a · B/B · c/c · D/D · e/e*.

O geneticista cruza a F_1 desse cruzamento com um testador recessivo e obtém:

A · B · C · D · E	243
a · B · c · D · e	237

$A \cdot B \cdot c \cdot D \cdot e$	62
$a \cdot B \cdot C \cdot D \cdot E$	58
$A \cdot B \cdot C \cdot D \cdot e$	155
$a \cdot B \cdot c \cdot D \cdot E$	165
$a \cdot B \cdot C \cdot D \cdot e$	46
$A \cdot B \cdot c \cdot D \cdot E$	34
	1.000

O geneticista também sabe que os genes *D* e *E* se distribuem de modo independente.

a. Desenhe um mapa desses genes, demonstrando as distâncias em unidades de mapa sempre que possível.

b. Existe qualquer evidência de interferência?

65. Na planta *Arabidopsis*, os *loci* em relação ao comprimento da vagem (*L* = Longa; *l* = Curta) e aos pelos do fruto (*H* = Piloso; *h* = Liso) estão ligados a uma distância de 16 u.m. no mesmo cromossomo. Foram realizados os cruzamentos a seguir:

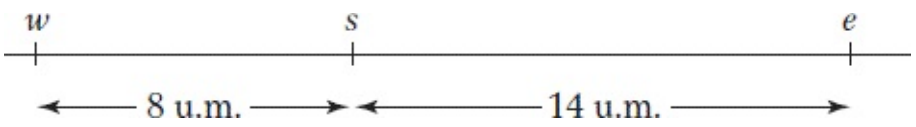
(I) *L H/L H* × *l h/l h* → F_1
(II) *L h/L h* × *l H/l H* → F_1

Se as F_1 dos cruzamentos I e II forem cruzadas:

a. Qual proporção da progênie espera-se que seja *l h/ l h*?

b. Qual proporção da progênie espera-se que seja *L h/ l h*?

66. No milho (*Zea mays*), o mapa genético de parte do cromossomo 4 é como segue, onde *w*, *s* e *e* representam alelos mutantes recessivos que afetam a cor e o formato do pólen:



Se o cruzamento a seguir for realizado:

+ + +/+ + + × *w s e*/*w s e*

e a F_1 for submetida ao cruzamento-teste com *w s e*/ *w s e* e se for presumido que não há interferência nessa região do cromossomo, qual proporção da progênie será dos genótipos a seguir?

a. + + + **e.** + + *e*

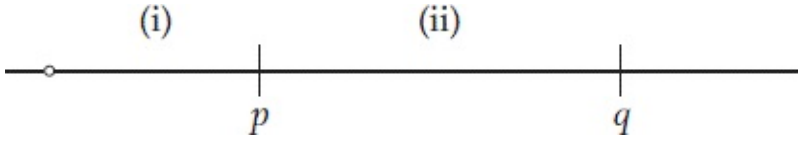
b. *w s e* **f.** *w s* +

c. + *s e* **g.** *w* + *e*

d. *w* + + **h.** + *s* +

67. Toda sexta-feira à noite, a estudante de genética Jean Allele, exausta em virtude de seus estudos, vai ao boliche da reunião de estudantes para relaxar. Entretanto, até lá, ela é assombrada por seus estudos genéticos. A pista de boliche um tanto quanto modesta apresenta apenas quatro bolas de boliche: duas vermelhas e duas azuis. Elas são jogadas contra os pinos e em seguida são coletadas e devolvidas pela canaleta em ordem aleatória, repousando na parada final. Na medida em que a noite passa, Jean observa padrões familiares das quatro bolas na medida em que elas chegam até o repouso na parada. Compulsivamente, ela conta os diferentes padrões. Quais padrões ela visualizou, quais foram as suas frequências e qual é a relevância desse assunto para a genética?

68. Em uma análise de tétrade, o arranjo de ligação dos *loci p* e *q* é como segue:



Presuma que:

- Na região I, não há *crossover* em 88% das meioses e há um *crossover* único em 12% das meioses
- Na região II, não há *crossover* em 80% das meioses e há um *crossover* único em 20% das meioses
- Não há interferência (em outras palavras, a situação em uma região não

afeta o que está ocorrendo na outra região).

Quais proporções de tétrades serão dos tipos a seguir? **(a)** M_IM_I, DP; **(b)** M_IM_I, DNP; **(c)** M_IM_{II}, T; **(d)** $M_{II}M_I$, T; **(e)** $M_{II}M_{II}$, DP; **(f)** $M_{II}M_{II}$, DNP; **(g)** $M_{II}M_{II}$, T. (Nota: aqui, o padrão M escrito primeiramente é aquele que se refere ao *locus p*.) **Dica**: o modo mais fácil de resolver esse problema é iniciar com o cálculo das frequências de ascos com *crossovers* em ambas as regiões, na região I, na região II e em nenhuma região. Em seguida determine quais padrões de M_I e M_{II} resultam.

69. Para um experimento com uma levedura haploide, você tem duas culturas diferentes. Cada uma será cultivada em meio mínimo ao qual foi adicionada arginina, mas nenhuma será cultivada apenas em meio mínimo. (O meio mínimo é composto por sais inorgânicos mais açúcar.) Com a utilização dos métodos apropriados, você induz as duas culturas ao cruzamento. As células diploides em seguida se dividem por meiose e formam tétrades não ordenadas. Alguns dos ascósporos serão cultivados em meio mínimo. Você classifica um grande número dessas tétrades em relação aos fenótipos ARG$^-$ (necessidade de arginina) e ARG$^+$ (independente de arginina) e registra os dados a seguir:

Segregação de ARG$^-$:ARG$^+$	Frequência (%)
4:0	40
3:1	20
2:2	40

a. Com a utilização de símbolos de sua própria escolha, atribua os genótipos para as duas culturas parentais. Em relação a cada um dos três tipos de segregação, atribua genótipos aos segregantes.

b. Se houver mais de um *locus* controlando a necessidade de arginina, esses

loci estão ligados?

70. Uma análise de RFLP de duas linhagens puras *A/A · B/B* e *a/a · b/b* demonstrou que a primeira era homozigota em relação a um alelo de RFLP longo (*l*) e que a última, em relação a um alelo curto (*s*). As duas foram cruzadas para formar uma F_1, que em seguida foi retrocruzada com a segunda linhagem pura. Uma progênie de mil foi classificada como segue:

Aa BB ss	9	*Aa bb ss*	43
Aa Bb ls	362	*Aa bb ls*	93
aa bb ls	11	*aa Bb ls*	37
aa bb ss	358	*aa Bb ss*	87

a. O que esses resultados nos informam a respeito de ligação?

b. Desenhe um mapa, se apropriado.

c. Incorpore os fragmentos de RFLP no seu mapa.

CAPÍTULO 5

A Genética das Bactérias e Seus Vírus

Bactérias em divisão. (Custom Medical Stock Photo RM/Getty Images.)

TÓPICOS

5.1 Trabalho com microrganismos
5.2 Conjugação bacteriana
5.3 Transformação bacteriana

5.4 Genética de bacteriófagos
5.5 Transdução
5.6 Comparação de mapas físicos e mapas de ligação

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Distinguir entre os procedimentos experimentais e as análises nos três principais modos por meio dos quais bactérias trocam genes
- Mapear os genomas bacterianos com a utilização da conjugação interrompida
- Mapear os genomas bacterianos com a utilização da frequência de recombinantes
- Avaliar o desfecho de experimentos de transformação dupla em termos de ligação
- Prever os desfechos de experimentos de transdução com a utilização de fagos capazes de transdução generalizada ou restrita
- Mapear os genomas dos fagos por meio da recombinação em infecções duplas de bactérias
- Desenhar experimentos para mapear uma mutação causada por mutagênese por transpóson
- Prever a herança dos genes e das funções contidas em plasmídios em cruzamentos bacterianos.

A tecnologia de DNA é responsável pelos rápidos avanços que estão sendo realizados na genética de todos os organismos-modelo. Ela também é um tópico de interesse considerável no domínio público. Exemplos são os anúncios intensamente publicados das sequências completas dos genomas de seres humanos e chimpanzés nos últimos anos e a popularidade da análise forense com base no DNA em séries de televisão e filmes (Figura 5.1). De fato, melhoras na tecnologia levaram ao sequenciamento dos genomas de muitas centenas de espécies. Os referidos resultados dramáticos, sejam em seres humanos, peixes,

insetos, plantas ou fungos, são todos baseados na utilização de métodos que possibilitam que pequenas partes do DNA sejam isoladas, carreadas de uma célula para outra e amplificadas em grandes amostras puras. Quase todos os sistemas sofisticados que possibilitam essas manipulações do DNA de qualquer organismo são derivados de bactérias e seus vírus. Assim sendo, o avanço da genética moderna até o seu presente estado de compreensão foi totalmente dependente do desenvolvimento da genética bacteriana, o tópico deste capítulo.

Entretanto, o objetivo da genética bacteriana nunca foi facilitar a genética molecular eucariótica. As bactérias são biologicamente importantes por si próprias. Elas são os organismos mais numerosos em nosso planeta. Elas contribuem para a reciclagem de nutrientes tais como nitrogênio, enxofre e carbono nos ecossistemas. Algumas são agentes de doenças em seres humanos, animais e plantas. Outras vivem em simbiose dentro de nossas bocas e de nossos intestinos. Além disso, muitos tipos de bactérias são úteis para a síntese industrial de uma ampla diversidade de produtos orgânicos. Portanto, o ímpeto para a dissecção genética das bactérias tem sido o mesmo que aquele em relação aos organismos multicelulares — compreender a sua função biológica.

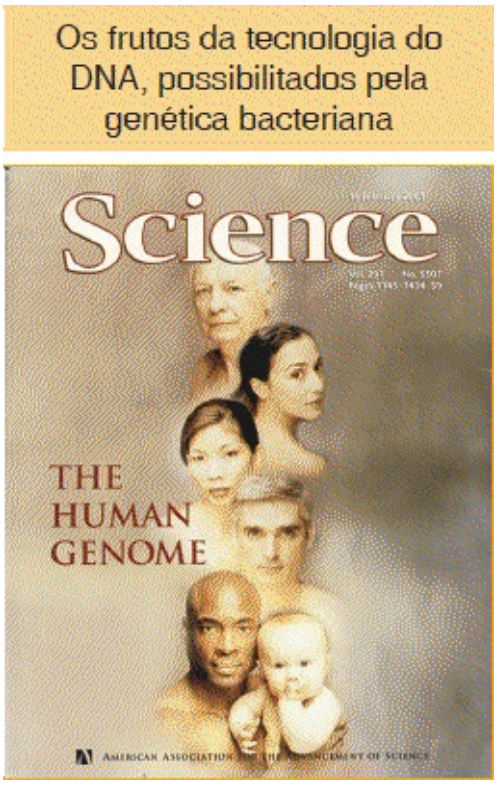


FIGURA 5.1 Os resultados dramáticos da tecnologia do DNA moderna, tal como o sequenciamento do genoma humano, foram possíveis apenas porque a genética bacteriana levou à invenção de eficientes vetores para manipulação do DNA. (*Science* 291, 2001, pp. 1145-1434. Imagem por Ann E. Cutting. Reproduzida, com autorização, da AAAS.)

As bactérias pertencem a uma classe de organismos conhecidos como **procariotos**, que também incluem algas azuis e verdes (classificadas como *cianobactérias*). Uma característica-chave da definição dos procariotos é que o seu DNA não está contido em um núcleo limitado por uma membrana. Assim como os organismos superiores, as bactérias possuem genes compostos por DNA dispostos em uma longa série em um "cromossomo". Entretanto, a organização do seu material genético é única em diversos aspectos. O genoma da maior parte das bactérias é uma molécula única de DNA com filamento duplo no formato de um círculo fechado. Além disso, as bactérias na natureza com frequência contêm elementos extras de DNA, denominados plasmídios. A maior parte dos plasmídios também são círculos de DNA, mas muito menores do que o genoma bacteriano principal.

As bactérias podem ser parasitadas por **vírus** específicos denominados

bacteriófagos, ou simplesmente **fagos**. Os fagos e outros vírus são muito diferentes dos organismos que temos estudado até agora. Os vírus apresentam algumas propriedades em comum com os organismos. Por exemplo, seu material genético pode ser DNA ou RNA, constituindo um "cromossomo" curto. Entretanto, a maior parte dos biólogos considera os vírus como não vivos, tendo em vista que eles não são células e não apresentam um metabolismo próprio. Portanto, para estudar sua genética, os vírus precisam ser propagados nas células de seus organismos hospedeiros.

Quando os cientistas começaram a estudar as bactérias e os fagos, eles estavam naturalmente curiosos a respeito dos seus sistemas hereditários. Claramente, as bactérias e os fagos têm sistemas hereditários, uma vez que demonstram aspecto e função constantes de uma geração para a próxima (realmente são constantes). Mas como esses sistemas hereditários atuam? As bactérias, assim como os organismos eucarióticos unicelulares, reproduzem-se assexuadamente por meio do crescimento e da divisão celular, uma célula se tornando duas. Essa reprodução assexuada é consideravelmente fácil de ser demonstrada experimentalmente. Entretanto, chega a ocorrer alguma união de tipos diferentes para fins da reprodução sexuada? Além disso, como os fagos muito menores se reproduzem? Eles chegam a se unir para um ciclo semelhante ao sexuado? Essas questões são buscadas neste capítulo.

Veremos que existe uma diversidade de processos hereditários em bactérias e fagos. Esses processos são interessantes em virtude da biologia básica dessas formas, mas também atuam como modelos — como fontes de percepção dos processos genéticos em atuação em *todos* os organismos. Para um geneticista, o atrativo dessas formas é que elas podem ser cultivadas em números muito grandes em virtude de serem tão pequenas. Consequentemente, é possível detectar e estudar *eventos genéticos* muito *raros* que são difíceis ou impossíveis de estudar em eucariotos.

Quais processos hereditários são observados em procariotos? Eles podem apresentar tanto a reprodução assexuada quanto a sexuada. A mutação ocorre em células assexuadas em grande parte do mesmo modo como ela ocorre em eucariotos e os alelos mutantes podem ser seguidos por meio de ambos esses

processos em uma abordagem análoga àquela utilizada em eucariotos. Seguiremos os alelos desse modo no capítulo adiante.

Quando as células bacterianas se reproduzem assexuadamente, seu DNA genômico se replica e é repartido para as células-filhas, mas o método de partição é consideravelmente diferente da mitose.

Na reprodução sexuada, duas moléculas de DNA de diferentes fontes são reunidas. Entretanto, uma diferença importante dos eucariotos é que, nas bactérias, raramente dois cromossomos completos são reunidos; normalmente, a união é de um cromossomo completo com um fragmento de outro. As possibilidades estão resumidas na Figura 5.2.

O primeiro processo de troca de genes a ser examinado será a *conjugação*, que é o contato e a fusão de duas células bacterianas diferentes. Após a fusão, uma célula, denominada doadora, por vezes transfere o DNA genômico para a outra célula. Esse DNA transferido pode ser parte do ou (raramente) todo o genoma bacteriano. Em alguns casos, um ou mais elementos de DNA extragenômico autônomos, denominados plasmídios, se estiverem presentes, são transferidos. Os referidos plasmídios conseguem carrear o DNA genômico para dentro da célula receptora. Qualquer fragmento genômico transferido por meio de qualquer via pode se recombinar com o cromossomo da receptora após a entrada.

Uma célula bacteriana também pode captar um fragmento de DNA do ambiente externo e incorporar esse DNA ao seu próprio cromossomo, um processo denominado *transformação*. Além disso, determinados fagos podem captar um fragmento de DNA de uma célula bacteriana e injetá-lo em outra, onde ele pode ser incorporado ao cromossomo, em um processo conhecido como *transdução*.

A transferência do DNA em um plasmídio, por meio da transformação ou por meio da transdução, constitui um processo conhecido como **transmissão horizontal**, um tipo de transmissão gênica sem a necessidade de divisão celular. Esse termo distingue esse tipo de transferência de DNA daquele durante a **transmissão vertical**, a passagem do DNA entre as gerações bacterianas. A transmissão horizontal pode difundir o DNA rapidamente em uma população bacteriana por meio do contato, em grande parte do mesmo modo como uma doença se difunde. Para as bactérias, a transmissão horizontal proporciona um

método poderoso, por meio do qual elas conseguem se adaptar rapidamente às alterações nas condições ambientais.

Os próprios fagos podem sofrer recombinação quando dois genótipos diferentes infectam a mesma célula bacteriana (**recombinação de fagos**, não demonstrada na Figura 5.2).

Antes de analisarmos esses modos de troca genética, consideremos os modos práticos de manuseio de bactérias, que são muito diferentes daqueles utilizados no manuseio de organismos multicelulares.

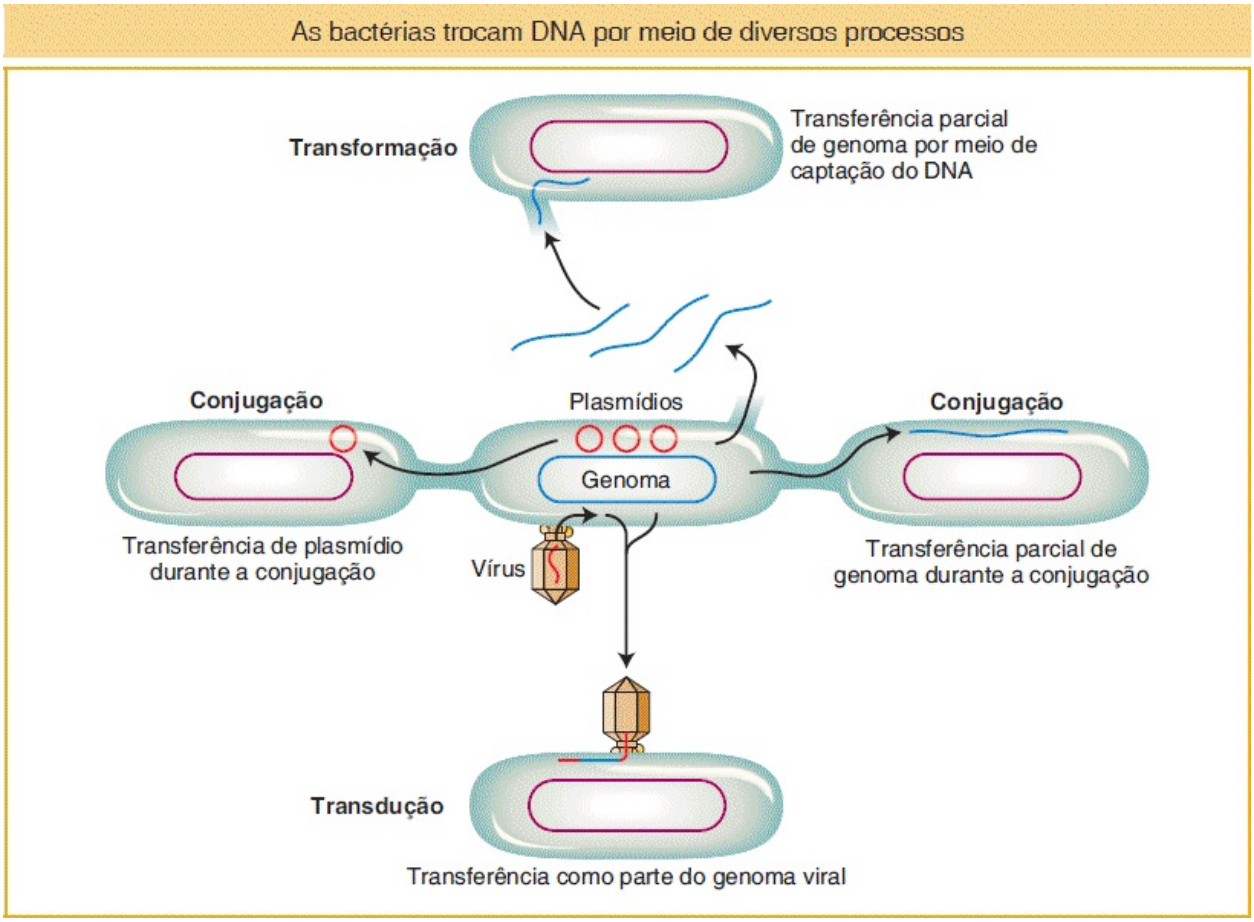


FIGURA 5.2 O DNA bacteriano pode ser transferido de uma célula para outra de quatro modos: conjugação com transferência de plasmídio, conjugação com transferência parcial de genoma, transformação e transdução.

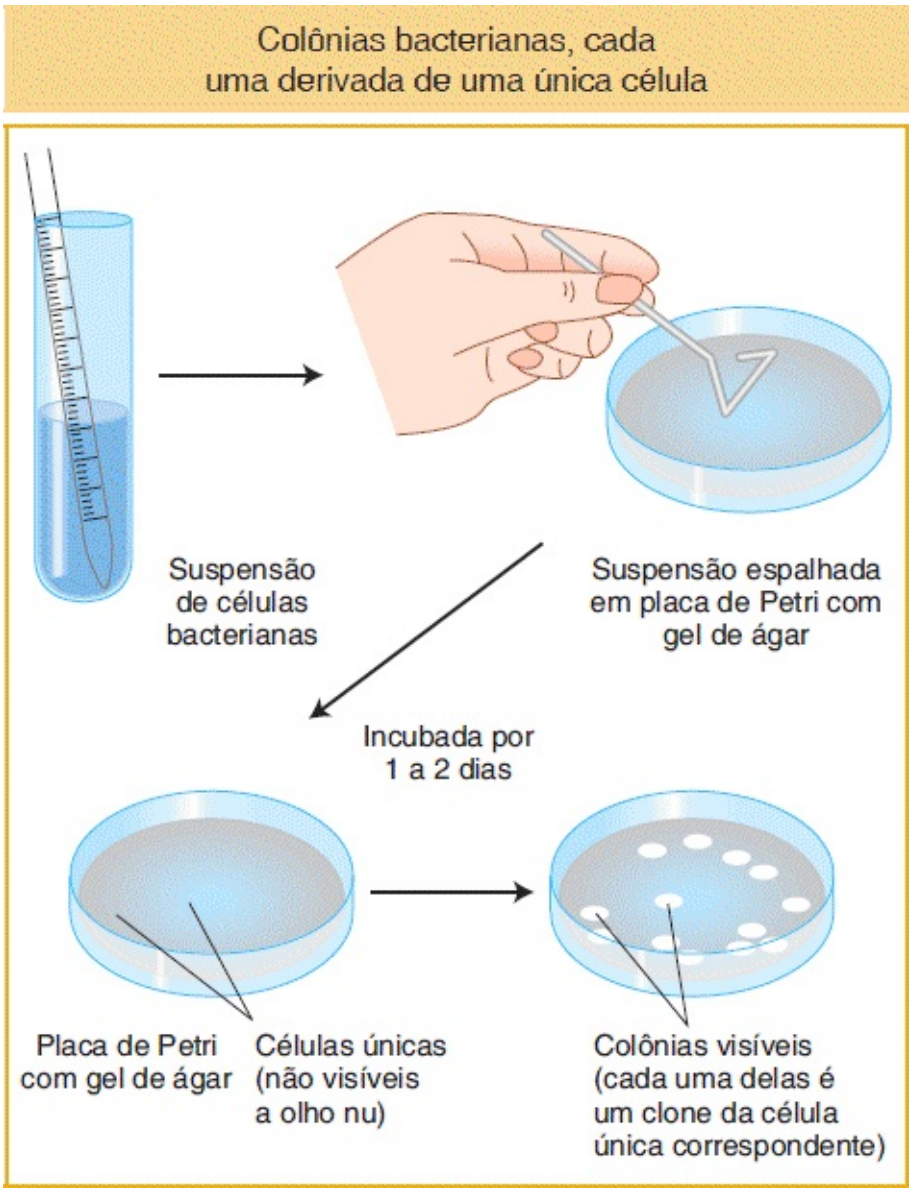


FIGURA 5.3 Fenótipos bacterianos podem ser avaliados em suas colônias. Um estoque de células bacterianas pode ser cultivado em um meio líquido que contenha nutrientes e em seguida um pequeno número de bactérias da suspensão líquida pode ser espalhado sobre um meio de ágar sólido. Cada uma das células dará origem a uma colônia. Todas as células em uma colônia apresentam os mesmos genótipo e fenótipo.

5.1 Trabalho com microrganismos

As bactérias têm divisão rápida e ocupam pouco espaço; assim, elas são muito convenientes como organismos-modelo genéticos. Elas podem ser cultivadas em um meio líquido ou em uma superfície sólida, como um gel de ágar, desde que os

nutrientes básicos sejam fornecidos. Cada célula bacteriana se divide assexuadamente de 1 → 2 → 4 → 8 → 16 células e assim por diante, até que os nutrientes se esgotem ou as escórias metabólicas tóxicas se acumulem em níveis que interrompam o crescimento da população. Uma pequena quantidade de uma cultura líquida pode ser pipetada sobre uma placa de Petri contendo meio de ágar sólido e espalhada de modo uniforme sobre a superfície com um difusor estéril, em um processo denominado **plaqueamento** (Figura 5.3). As células se dividem, mas, tendo em vista que não conseguem se dispersar na superfície do gel, todas as células permanecem agrupadas. Quando essa massa alcança mais de 10^7 células, ela se torna visível a olho nu como uma **colônia**. Cada colônia distinta na placa foi derivada de uma célula original única. Os membros de uma colônia que apresentam um ancestral genético único são conhecidos como **clones celulares**.

Os mutantes bacterianos são de obtenção consideravelmente fácil. Mutantes nutricionais são um bom exemplo. Bactérias do tipo selvagem são **prototróficas**, o que significa que podem crescer e se dividir em **meio mínimo** — um substrato que contém apenas sais inorgânicos, uma fonte de carbono para energia e água. A partir de uma cultura prototrópica, podem ser obtidos mutantes **auxotróficos**: esses mutantes são células que não crescerão, exceto se o meio contiver um ou mais blocos de construção celular específicos, tais como adenina, treonina ou biotina. Outro tipo de mutante útil difere do tipo selvagem na capacidade de utilizar uma fonte de energia específica; por exemplo, o tipo selvagem (lac^+) consegue utilizar a lactose e crescer, enquanto um mutante (lac^-) não consegue. A Figura 5.4 demonstra outro modo para distinguir as colônias lac^+ e lac^- por meio da utilização de um corante. Em outra categoria mutante, enquanto os tipos selvagens são suscetíveis a um inibidor tal como o antibiótico estreptomicina, os **mutantes resistentes** conseguem se dividir e formar colônias na presença do inibidor. Todos esses tipos de mutantes possibilitam que o geneticista distinga as diferentes linhagens individuais, proporcionando, assim, **marcadores genéticos** (alelos marcadores) para seguir os genomas e as células em experimentos. A Tabela 5.1 resume alguns fenótipos bacterianos mutantes e seus símbolos genéticos.

As seções a seguir documentam a descoberta dos diversos processos por meio

dos quais os genomas bacterianos se recombinam. Os métodos históricos são interessantes por si próprios, mas também servem para introduzir os diversos processos de recombinação, bem como as técnicas analíticas que ainda são aplicáveis atualmente.

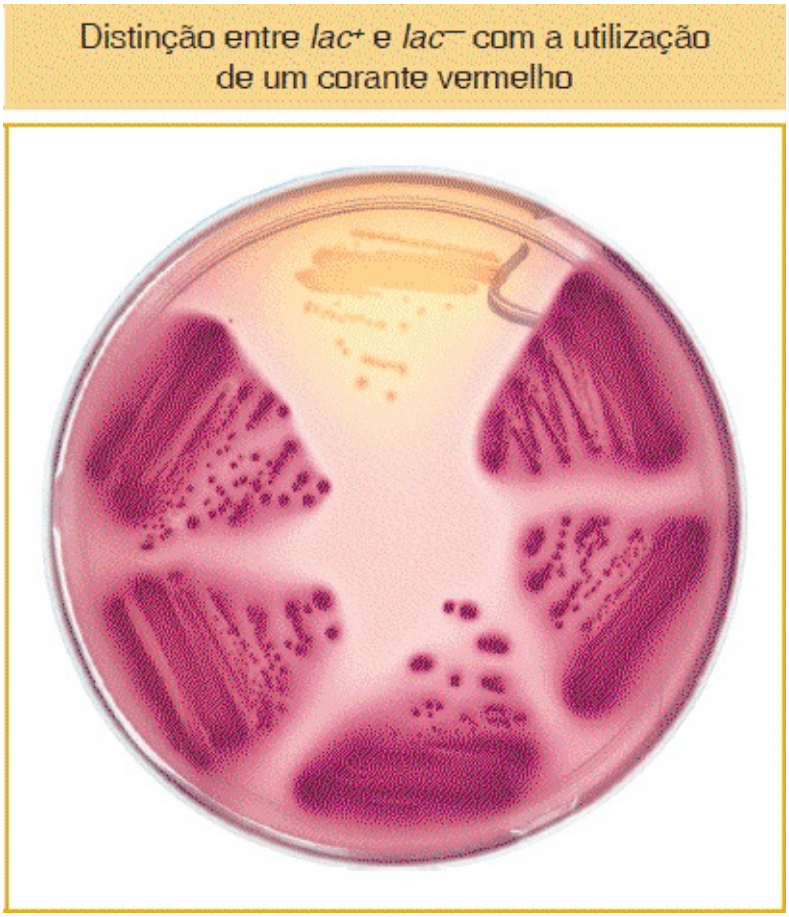


FIGURA 5.4 Bactérias do tipo selvagem capazes de utilizar a lactose como uma fonte de energia (lac^+) são coradas em vermelho na presença deste corante indicador. As células não coradas são mutantes incapazes de utilizar a lactose (lac^-). (Jeffrey H. Miller.)

Tabela 5.1 Alguns símbolos genotípicos utilizados em genética bacteriana.

Símbolo	Característica ou fenótipo associado ao símbolo
bio⁻	Requer a adição de biotina como um suplemento ao meio mínimo

arg^{-}	Requer a adição de arginina como um suplemento ao meio mínimo
met^{-}	Requer a adição de metionina como um suplemento ao meio mínimo
lac^{-}	Não consegue utilizar a lactose como uma fonte de carbono
gal^{-}	Não consegue utilizar a galactose como uma fonte de carbono
str^{r}	Resistente ao antibiótico estreptomicina
str^{s}	Sensível ao antibiótico estreptomicina

Nota: o meio mínimo é o meio sintético básico para o crescimento bacteriano sem suplementos de nutrientes.

5.2 Conjugação bacteriana

Os estudos mais iniciais em genética bacteriana revelaram o processo inesperado da conjugação celular.

Descoberta da conjugação

As bactérias apresentam um processo semelhante à reprodução sexuada e à recombinação? A questão foi respondida pelo trabalho experimental elegantemente simples de Joshua Lederberg e Edward Tatum, que em 1946 descobriram um processo semelhante ao sexuado no que se tornou o principal

modelo para a genética bacteriana, a *Escherichia coli* (ver Organismo-modelo, adiante). Eles estavam estudando duas linhagens de *E. coli* com diferentes conjuntos de mutações auxotróficas. A linhagem A⁻ cresceria apenas se o meio fosse suplementado com metionina e biotina; a linhagem B⁻ cresceria apenas se ele fosse suplementado com treonina, leucina e tiamina. Portanto, podemos designar as linhagens como:

Linhagem A⁻: *met*⁻ *bio*⁻ *thr*⁺ *leu*⁺ *thi*⁺
Linhagem B⁻: *met*⁺ *bio*⁺ *thr*⁻ *leu*⁻ *thi*⁻

A Figura 5.5 A ilustra de modo simplificado o desenho de seu experimento. As linhagens A⁻ e B⁻ foram misturadas em conjunto, incubadas durante um período e em seguida plaqueadas em meio mínimo, no qual nenhum auxotrófico poderia crescer. Observou-se que uma minoria das células (1 em 10^7) crescia como prototróficos e, portanto, necessariamente era do tipo selvagem, tendo recuperado a capacidade de crescer sem a adição de nutrientes. Algumas das placas foram plaqueadas apenas com bactérias da linhagem A⁻ e algumas com apenas bactérias da linhagem B⁻ para atuar como controles, mas nenhum prototrófico surgiu a partir desses plaqueamentos. A Figura 5.5 B ilustra o experimento em mais detalhes. Esses resultados sugeriram que algum tipo de recombinação de genes havia ocorrido entre os genomas das duas linhagens para produzir os prototróficos.

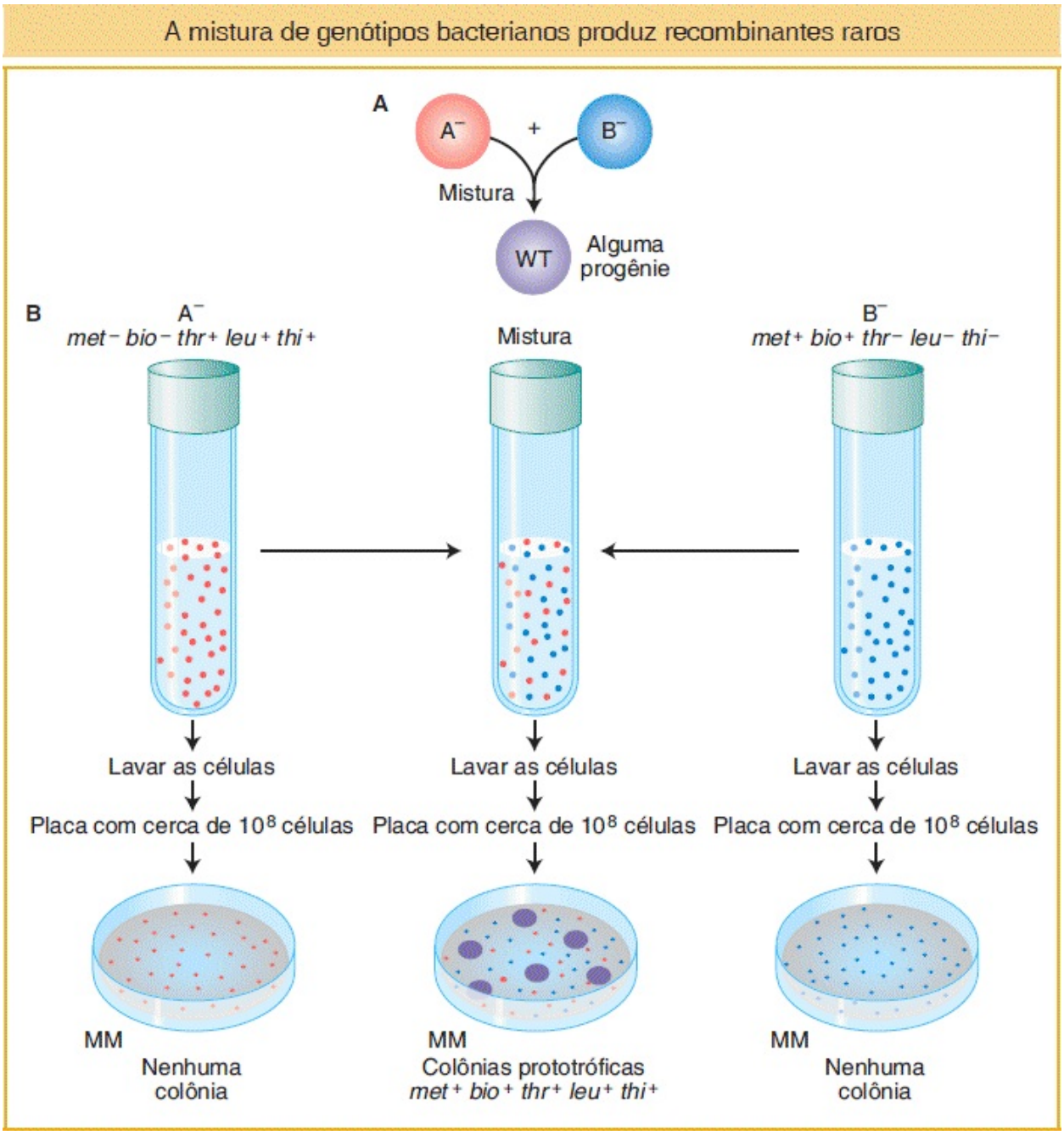


FIGURA 5.5 Com a utilização deste método, Lederberg e Tatum demonstraram que a recombinação genética entre genótipos bacterianos é possível. **A.** O conceito básico: duas culturas auxotróficas (A^- e B^-) são misturadas, produzindo tipos selvagens (WT) prototróficos. **B.** Células do tipo A^- ou do tipo B^- não podem ser cultivadas em um meio não suplementado (mínimo) (MM), tendo em vista que cada uma de A^- e B^- carreia mutações que causam a incapacidade de sintetizar os constituintes necessários para o crescimento celular. Entretanto, quando A^- e B^- são misturadas durante algumas horas e em seguida são plaqueadas, algumas poucas colônias aparecem na placa de ágar. Essas colônias derivam de células únicas nas quais o material genético foi trocado. Assim, são capazes de sintetizar todos os constituintes necessários do

metabolismo.

Poderia ser argumentado que as células das duas linhagens na realidade não trocam genes, mas, em vez disso, extravasam substâncias que as outras células conseguem absorver e utilizar para crescimento. Essa possibilidade de "alimentação cruzada" foi descartada por Bernard Davis como segue. Ele construiu um tubo com formato de U, no qual os dois braços eram separados por um filtro fino. Os poros do filtro eram muito pequenos para possibilitar a passagem das bactérias, mas grandes o suficiente para possibilitar a fácil passagem de quaisquer substâncias dissolvidas (Figura 5.6). A linhagem A^- foi colocada em um braço, a linhagem B^- no outro. Após as linhagens terem sido incubadas durante um período, Davis testou o conteúdo de cada braço para verificar se havia quaisquer células prototróficas, mas nenhuma foi observada. Em outras palavras, era necessário o *contato físico* entre as duas linhagens para a formação das células do tipo selvagem. Aparentemente havia ocorrido algum tipo de união de genoma e recombinantes genuínos haviam sido produzidos. A união física das células bacterianas pode ser confirmada sob um microscópio eletrônico e atualmente é denominada **conjugação** (Figura 5.7).

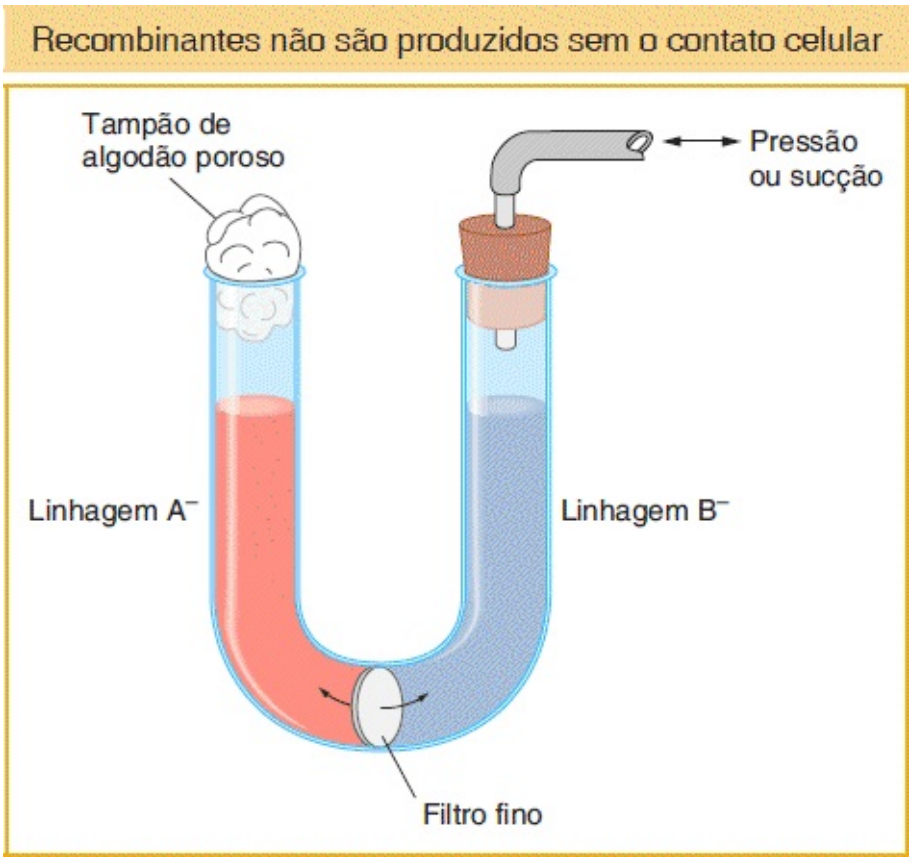


FIGURA 5.6 Linhagens bacterianas auxotróficas A^{-} e B^{-} são cultivadas em qualquer um dos lados de um tubo com formato de U. O líquido pode passar entre os braços por meio da aplicação de pressão ou sucção, mas as células bacterianas não conseguem passar pelo filtro. Após a incubação e o plaqueamento, nenhuma colônia recombinante cresce em meio mínimo.

FIGURA 5.7 Uma célula doadora estende uma ou mais projeções, ou *pili*, que se unem a uma célula receptora e aproximam as duas bactérias. (Dr. L. Caro/Science Source.)

Descoberta do fator de fertilidade (F)

Em 1953, William Hayes descobriu que, nos tipos de "cruzamentos" há pouco descritos aqui, os genitores conjugantes atuavam *de modo desigual* (veremos modos para demonstrar essa participação desigual posteriormente). Um genitor (e *apenas* esse genitor) aparentava transferir uma parte do ou todo o seu genoma para a outra célula. Portanto, uma célula atua como **doadora** e a outra célula atua como uma **receptora**. Esse "cruzamento" é consideravelmente diferente dos cruzamentos eucarióticos, nos quais os genitores contribuem igualmente com genomas nucleares para um indivíduo da progênie.

> CONCEITO-CHAVE A transferência do material genético na conjugação de *E. coli* não é recíproca. Uma célula, a doadora, transfere parte de seu genoma para outra célula, que atua como receptora.

Acidentalmente, Hayes descobriu uma variante de sua linhagem doadora original que não produzia recombinantes no cruzamento com a linhagem receptora. Aparentemente, a linhagem do tipo doadora havia perdido a capacidade de transferir o material genético e havia sido alterada para uma linhagem do tipo receptora. Ao trabalhar com sua variante doadora "estéril", Hayes observou que ela podia recuperar a capacidade de atuar como doadora por meio da associação com outras linhagens doadoras. De fato, a capacidade de doadora foi transmitida de modo rápido e efetivo entre as linhagens durante a conjugação. Um tipo de "transferência infecciosa" de algum fator aparentava estar ocorrendo. Ele sugeriu que a capacidade doadora por si própria é um estado hereditário, imposto por um **fator de fertilidade (F)**. As linhagens que carreiam o F conseguem doar e são denominadas $\mathbf{F^+}$. As linhagens com ausência do F não conseguem doar e são receptoras, denominadas $\mathbf{F^-}$.

Agora sabemos muito mais a respeito do F. Ele é um exemplo de uma molécula

de DNA circular pequena e não essencial, denominada **plasmídio**, que consegue se replicar no citoplasma independentemente do cromossomo do hospedeiro. A Figura 5.8 demonstra como as bactérias conseguem transferir plasmídios tal como o F. O plasmídio F orienta a síntese dos *pili* (singular, *pilus*), projeções que iniciam o contato com uma receptora (ver Figuras 5.7 e 5.8) e que a puxam para mais perto. O DNA de F na célula doadora faz uma versão de si próprio de filamento único por um mecanismo peculiar, denominado **replicação por círculo rolante**. O plasmídio circular "rola" e na medida em que gira, desenrola o filamento único, como uma "linha de pesca". Esse filamento único passa por um poro para dentro da célula receptora, onde o outro filamento é sintetizado, formando uma dupla-hélice. Portanto, uma cópia de F permanece na doadora e outra aparece na receptora, conforme demonstrado na Figura 5.8. Observe que o genoma da *E. coli* é ilustrado como um cromossomo circular único na Figura 5.8. (Examinaremos a evidência em relação a isso posteriormente.) A maior parte dos genomas bacterianos é circular, uma característica consideravelmente diferente dos cromossomos nucleares eucarióticos. Veremos que essa característica leva às muitas idiossincrasias da genética bacteriana.

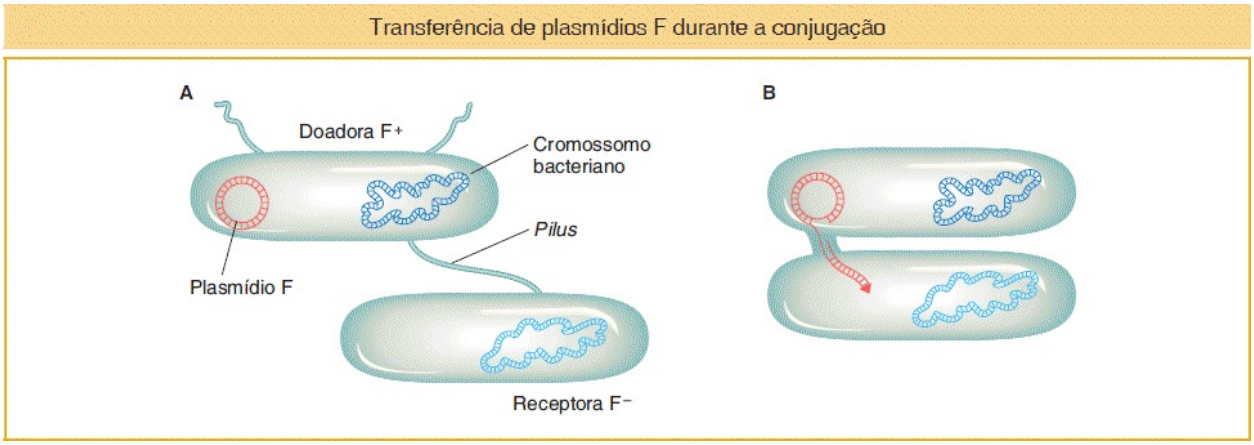


FIGURA 5.8 A. Durante a conjugação, o *pilus* aproxima duas bactérias. **B.** Em seguida, forma-se uma ponte entre as duas células. Uma cópia de filamento único do DNA do plasmídio é produzida na célula doadora e em seguida passa para a bactéria receptora, na qual o filamento único, atuando como um molde, é convertido na dupla-hélice.

Linhagens Hfr

Uma importante descoberta surgiu quando Luca Cavalli-Sforza descobriu um derivado de uma linhagem F^+ com duas propriedades incomuns:

1. No cruzamento com linhagens F^-, essa nova linhagem produziu 1.000 vezes mais recombinantes do que uma linhagem F^+ normal. Cavalli-Sforza designou essa derivada uma linhagem **Hfr** para simbolizar a sua capacidade de promover uma alta *f*requência de *r*ecombinação.
2. Em cruzamentos Hfr × F^-, virtualmente nenhum dos genitores F^- foi convertido em F^+ ou em Hfr. Esse resultado contrasta com os cruzamentos F^+ × F^-, nos quais, conforme vimos, a transferência infecciosa do F resulta em uma grande proporção de genitores F^- que são convertidos em F^+.

Ficou aparente que uma linhagem Hfr resulta da integração do fator F no cromossomo, conforme ilustrado na Figura 5.9. Agora podemos explicar a primeira propriedade incomum das linhagens Hfr. Durante a conjugação, o fator F inserido no cromossomo direciona com eficiência parte desse ou todo esse cromossomo para dentro da célula F^-. O fragmento cromossômico em seguida pode participar da recombinação com o cromossomo receptor. Os raros recombinantes observados por Lederberg e Tatum em cruzamentos F^+ × F^- ocorreram em virtude da formação espontânea, porém rara, de células Hfr na cultura F^+. Cavalli-Sforza isolou exemplos dessas células raras de culturas F^+ e observou que, de fato, agora elas atuavam como verdadeiras Hfr.

Uma célula Hfr morre após doar seu material cromossômico para uma célula F^-? A resposta é não. Assim como o plasmídio F, o cromossomo Hfr se replica e transfere um único filamento para a célula F^- durante a conjugação. O fato que o DNA transferido é um filamento único pode ser demonstrado visualmente com a utilização de linhagens e anticorpos especiais, conforme demonstrado na Figura 5.10. A replicação do cromossomo assegura um cromossomo completo para a célula doadora após o cruzamento. O filamento transferido é convertido em uma dupla-hélice na célula receptora e os genes doadores podem se tornar incorporados no cromossomo da receptora por meio de *crossovers*, criando uma célula recombinante (Figura 5.11). Se não houver recombinação, os fragmentos de DNA transferidos são simplesmente perdidos na progressão da divisão celular.

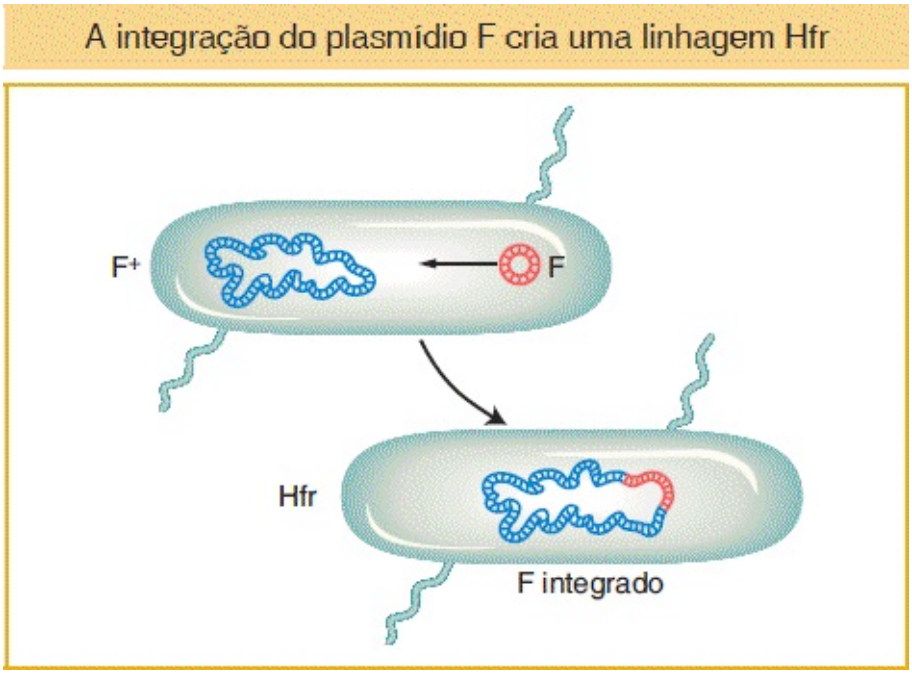


FIGURA 5.9 Em uma linhagem F^+, o plasmídio F livre ocasionalmente se integra ao cromossomo de *E. coli*, criando uma linhagem Hfr.

ORGANISMO-MODELO *Escherichia coli*

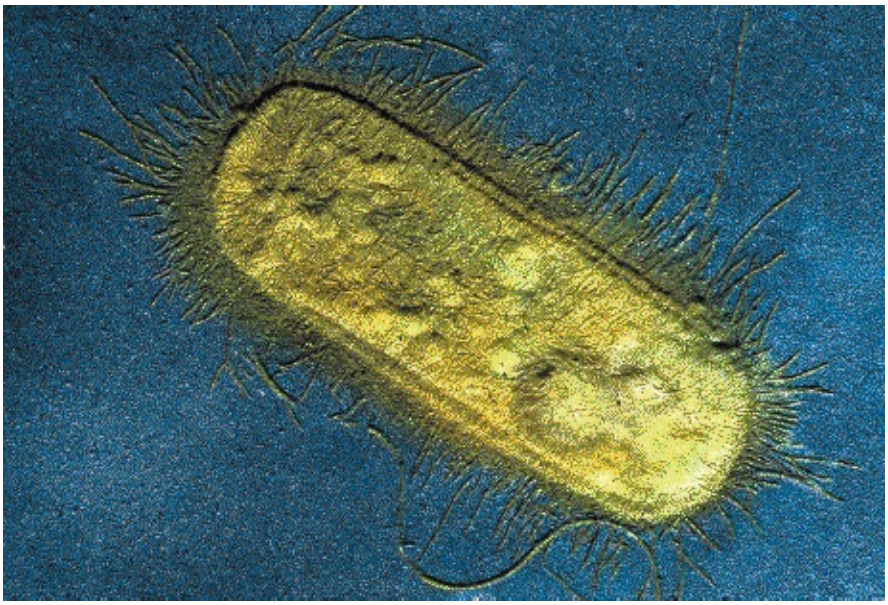
O microscopista do século 17 Antony van Leeuwenhoek provavelmente foi o primeiro a visualizar células bacterianas e a reconhecer o seu pequeno tamanho: "Existe mais vida na espuma sobre os dentes na boca de um homem do que existem homens em todo o reino." Entretanto, a bacteriologia só iniciou de fato no século 19. Na década de 1940, Joshua Lederberg e Edward Tatum fizeram a descoberta que lançou a bacteriologia no campo da genética em desenvolvimento: descobriram que, em uma determinada bactéria, havia um tipo de ciclo sexual, incluindo um processo semelhante ao *crossing over*. O organismo que eles escolheram para esse experimento se tornou o modelo não apenas para a genética dos procariotos, mas, em um sentido, para toda a genética. O organismo era a *Escherichia coli*, uma bactéria denominada em homenagem ao seu descobridor, o bacteriologista alemão do século 19 Theodore Escherich.

A escolha da *E. coli* foi uma sorte, tendo em vista que ela comprovadamente

apresenta muitas características adequadas para as pesquisas genéticas, isso além de ser facilmente obtida, tendo em vista que vive no intestino de seres humanos e de outros animais. No intestino, é uma simbionte benigna, mas ocasionalmente causa infecções urinárias e diarreia.

E. coli tem um cromossomo circular único com comprimento de 4,6 Mb. De seus 4.000 genes sem íntrons, aproximadamente 35% são de função desconhecida. O ciclo sexual é possibilitado por meio da ação de um plasmídio extragenômico denominado F, que confere um tipo de "masculinidade". Outros plasmídios carreiam genes cujas funções equipam a célula para a vida em ambientes específicos, tais como genes de resistência a fármacos. Esses plasmídios foram adaptados como *vetores* gênicos, que são transportadores de genes que formam a base das transferências gênicas no centro da engenharia genética moderna.

A *E. coli* é unicelular e cresce por simples divisão celular. Em virtude de seu tamanho pequeno (comprimento de cerca de 1 μm), a *E. coli* pode ser cultivada em grandes números e submetida à seleção intensiva e à triagem em relação a eventos genéticos raros. Pesquisas com *E. coli* representam o início do esclarecimento da "caixa preta" em genética: por meio da seleção e da análise de mutantes, as funções do maquinário genético puderam ser deduzidas, ainda que fossem muito pequenas para serem visualizadas. Fenótipos tais como tamanho da colônia, resistência a fármacos, utilização de fonte de carbono e produção de corante colorido tomaram o lugar dos fenótipos visíveis da genética de eucariotos.



Uma micrografia eletrônica de uma célula de *E. coli* demonstrando flagelos longos, utilizados para a locomoção, e fímbrias, pelos proteicos que são importantes para a ancoragem das células nos tecidos animais. (Os *pili* sexuais não estão demonstrados nesta micrografia.) (Biophoto Associates/Science Photo Library.)

Transmissão linear dos genes Hfr a partir de um ponto fixo. Uma visão mais clara do comportamento das linhagens Hfr foi obtida em 1957, quando Elie Wollman e François Jacob investigaram o padrão de transmissão de genes Hfr para células F $^-$ durante um cruzamento. Eles cruzaram:

Hfr *azi*^{r} *ton*^{r} *lac*$^+$ *gal*$^+$ *str*^{s} × F$^-$ *azi*^{s} *ton*^{s} *lac*$^-$ *gal*$^-$ *str*^{r}

(Os sobrescritos “r” e “s” fazem referência a resistente e sensível, respectivamente.) Em ocasiões específicas após a mistura, eles removeram amostras, cada uma das quais foi colocada em um liquidificador de cozinha durante alguns segundos para separar os pares de células em cruzamento. Esse procedimento é denominado **cruzamento interrompido**. Em seguida, a amostra foi plaqueada em um meio contendo estreptomicina para matar as células

doadoras Hfr, que continham o alelo de sensibilidade *str*^{s}. As células *str*^{r} sobreviventes em seguida foram testadas em relação à presença de alelos do genoma Hfr doador. Qualquer célula *str*^{r} que continha um alelo doador deve ter participado na conjugação: as referidas células são denominadas **exconjugantes**. Os resultados estão inseridos em um gráfico na Figura 5.12 A, que demonstra o momento de entrada de cada alelo doador *azi*^{r}, *ton*^{r}, *lac*$^{+}$ e *gal*$^{+}$. A Figura 5.12 B ilustra a transferência dos alelos Hfr.

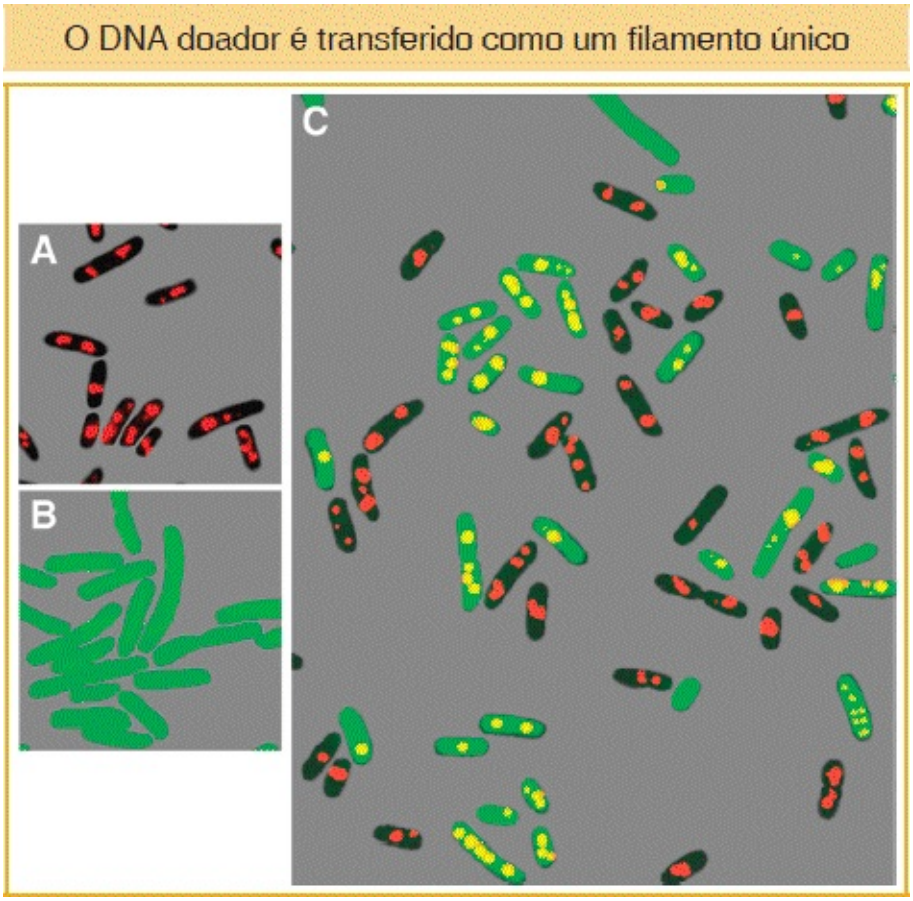


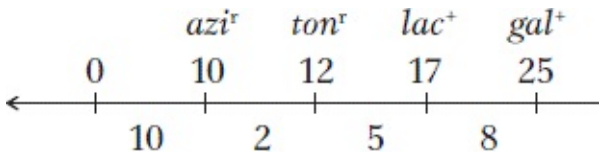
FIGURA 5.10 As fotografias demonstram uma visualização da transferência do DNA de filamento único na conjugação em células de *E. coli*, com a utilização de anticorpos fluorescentes especiais. As linhagens Hfr parentais (**A**) são pretas com DNA vermelho. O vermelho é da ligação de um anticorpo com uma proteína normalmente ligada ao DNA. As células receptoras F⁻ (**B**) são verdes em virtude da presença do gene para uma proteína de uma água-viva que fluoresce em verde, e, em virtude de serem mutantes em relação a um determinado gene, a proteína de seu DNA não se liga com os anticorpos. Quando o DNA de filamento único da doadora Hfr entra na receptora, ele promove a ligação atípica desta proteína, que fluoresce em amarelo neste fundo. A parte **C** demonstra Hfr (inalteradas) e exconjugantes (células que foram submetidas à

conjugação) com DNA transferido em amarelo. Algumas poucas células F^- não cruzadas estão visíveis. (De M. Kohiyama, S. Hiraga, I. Matic, and M. Radman, "Bacterial Sex: Playing Voyeurs 50 Years Later", *Science* 301, 2003, p. 803, Fig. 1. Reproduzida, com autorização, da AAAS.)

Os elementos-chave nesses resultados são:

1. Cada alelo doador aparece pela primeira vez nas receptoras F^- em uma ocasião específica após o início do cruzamento.
2. Os alelos doadores aparecem em uma sequência específica.
3. Os alelos doadores posteriores estão presentes em menos células receptoras.

Reunindo essas observações, Wollman e Jacob deduziram que, na conjugante Hfr, a transferência do DNA de filamento único tem início a partir de um ponto fixo no cromossomo doador, denominado **origem (O)**, e continua de modo linear. Atualmente sabe-se que o ponto O é o sítio no qual o plasmídio F é inserido. Quanto mais longe um gene estiver de O, mais tardiamente ele é transferido para F^-. O processo de transferência em geral será interrompido antes que os genes mais distantes sejam transferidos e, como resultado, esses genes são incluídos em menos exconjugantes. Observe que um tipo de mapa cromossômico pode ser produzido em unidades de minutos, com base na ocasião da entrada dos genes marcados. No exemplo na Figura 5.12, o mapa seria:



Como podemos explicar a segunda propriedade incomum dos cruzamentos Hfr, que os exconjugantes F^- raramente são convertidos em Hfr ou F^+? Quando Wollman e Jacob possibilitaram que os cruzamentos Hfr × F^- continuassem por até 2 horas antes da separação, eles observaram que de fato alguns poucos exconjugantes eram convertidos em Hfr. Em outras palavras, a parte de F que confere a capacidade doadora foi, por fim, transmitida, mas a uma frequência muito baixa. A raridade das exconjugantes Hfr sugeriu que o F inserido foi transmitido como o *último* elemento do cromossomo linear. Podemos resumir a

ordem de transmissão com o tipo de mapa geral a seguir, no qual a seta indica o sentido da transferência, que tem início com O:

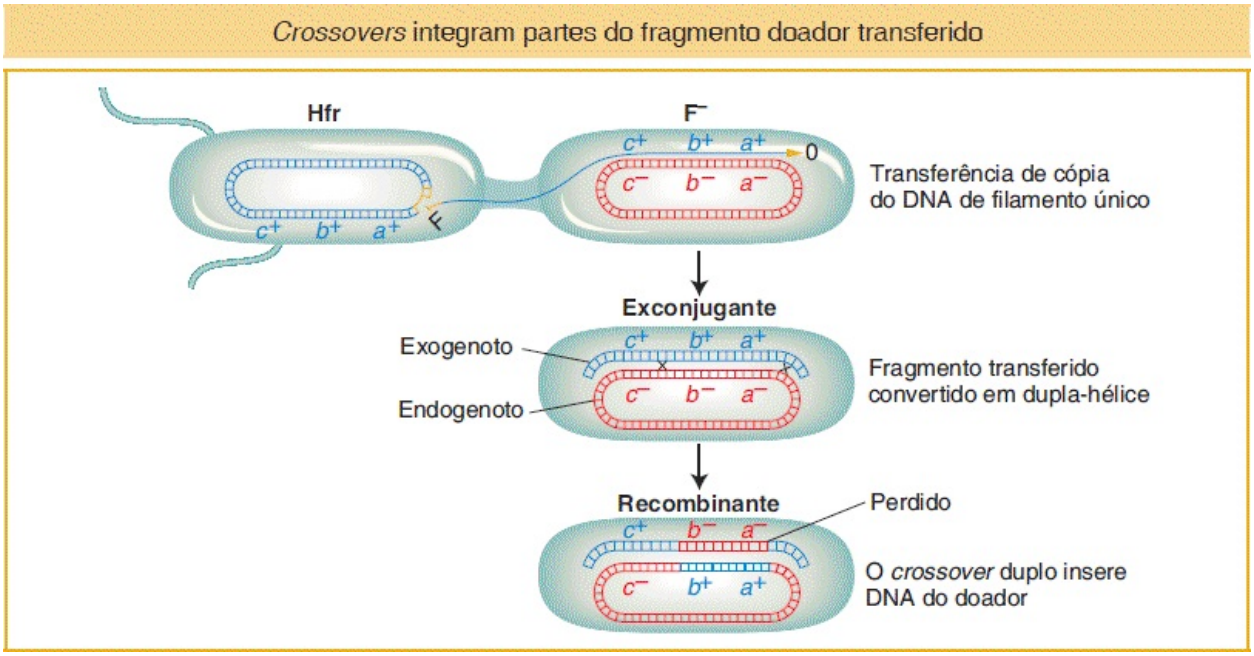


FIGURA 5.11 Após a conjugação, são necessários *crossovers* para integrar os genes do fragmento doador ao cromossomo receptor e, portanto, tornarem-se uma parte estável de seu genoma.

Acompanhamento do tempo de entrada do marcador gera um mapa cromossômico

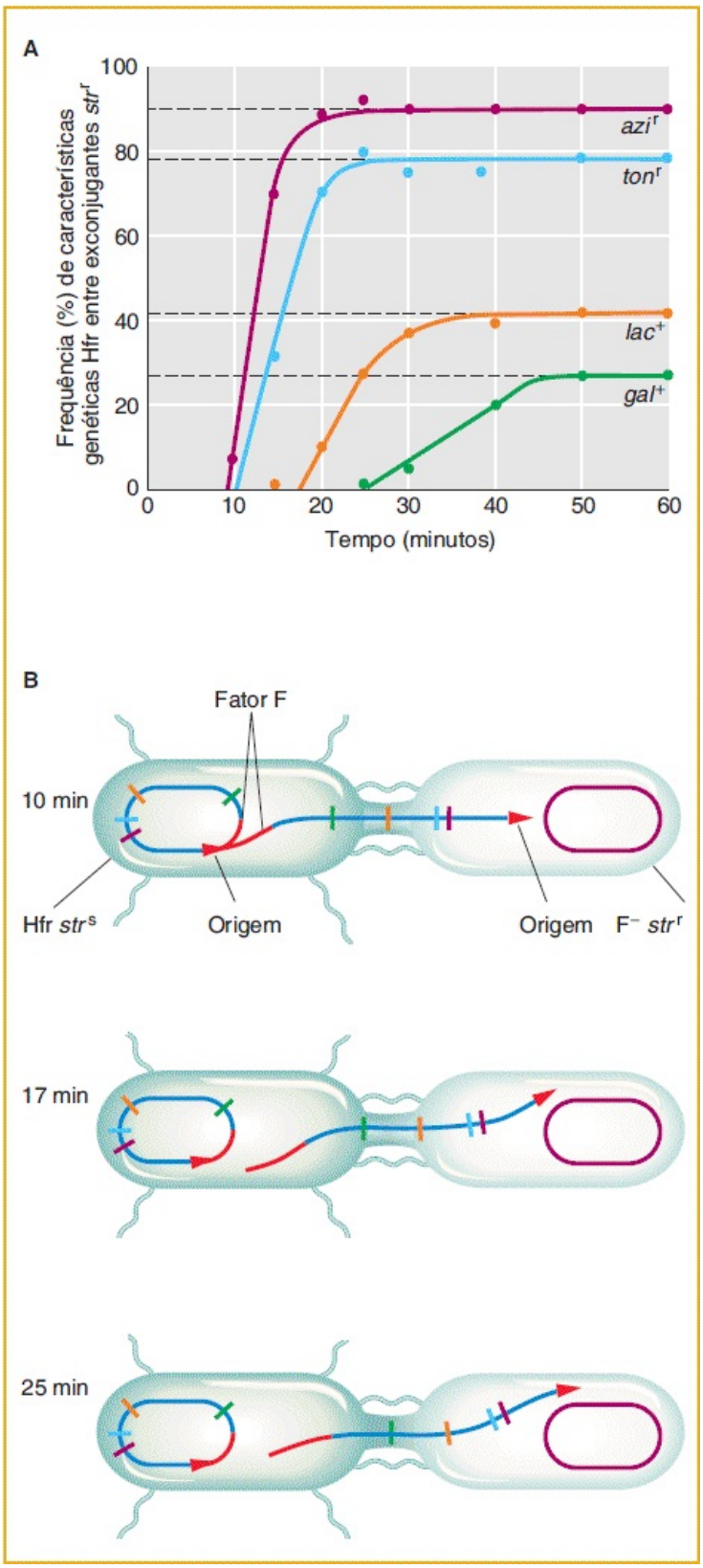
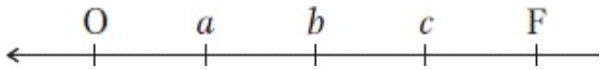


FIGURA 5.12 Neste experimento de conjugação com cruzamento interrompido, células F⁻ resistentes à estreptomicina com mutações em *azi, ton, lac* e *gal* são incubadas durante períodos variáveis com células Hfr que são sensíveis à estreptomicina e carreiam alelos do tipo selvagem em relação a esses genes. **A.** Um gráfico da frequência dos alelos doadores em exconjugantes em relação ao tempo após o cruzamento. **B.** Uma ilustração esquemática da transferência de marcadores (demonstrados em cores diferentes) com o passar do tempo. (*A.* Dados de E. L. Wollman, F. Jacob, and W. Hayes, Cold Spring Harbor Symp. *Quant. Biol.* 21, 1956, 141.)



Portanto, quase nenhuma das receptoras F⁻ é convertida, tendo em vista que o fator de fertilidade é o último elemento transferido e normalmente o processo de transmissão terá sido interrompido antes de chegar a esse ponto.

> CONCEITO-CHAVE O cromossomo Hfr, originalmente circular, desenrola uma cópia de si próprio, que é transferida para a célula F⁻ de modo linear, com o fator F entrando por último.

Inferência dos sítios de integração de F e a circularidade cromossômica. Wollman e Jacob prosseguiram para esclarecer melhor como e onde o plasmídio F se integra para formar uma célula Hfr e, ao fazer isso, deduziram que o cromossomo é circular. Eles realizaram experimentos de cruzamento interrompido com linhagens Hfr diferentes e derivadas em separado. Significativamente, a ordem de transmissão dos alelos diferiu de linhagem para linhagem, como nos exemplos a seguir:

Linhagem Hfr	
H	O *thr pro lac pur gal his gly thi* F
1	O *thr thi gly his gal pur lac pro* F

2	O *pro thr thi gly his gal pur lac* F
3	O *pur lac pro thr thi gly his gal* F
AB 312	O *thi thr pro lac pur gal his gly* F

Cada linhagem pode ser considerada um mapa que demonstra a ordem dos alelos no cromossomo. À primeira vista, aparentemente há uma mistura aleatória de genes. Entretanto, quando alguns dos mapas de Hfr são invertidos, a relação das sequências se torna clara.

H (escrita em ordem invertida)	F *thi gly his gal pur lac pro thr* O
1	O *thr thi gly his gal pur lac pro* F
2	O *pro thr thi gly his gal pur lac* F
3	O *pur lac pro thr thi gly his gal* F
AB 312 (escrita em ordem invertida)	F *gly his gal pur lac pro thr thi* O

A relação das sequências entre si é explicada se cada mapa for o segmento de um círculo. Essa foi a primeira indicação de que os cromossomos bacterianos são circulares. Além disso, Allan Campbell propôs uma hipótese surpreendente, que explicou os diferentes mapas de Hfr. Ele propôs que, se F é um anel, então a inserção poderia ocorrer por meio de um *crossover* simples entre F e o cromossomo bacteriano (Figura 5.13). Sendo esse o caso, quaisquer dos cromossomos lineares Hfr podem ser gerados simplesmente por meio da inserção de F no anel no local e na orientação apropriados (Figura 5.14).

Diversas hipóteses — amparadas posteriormente — se seguiram a partir da proposta de Campbell.

1. Uma extremidade do fator F integrado seria a origem, na qual tem início a transferência do cromossomo Hfr. O **término** estaria na outra extremidade de F.
2. A orientação na qual F é inserido determinaria a ordem de entrada dos alelos doadores. Se o círculo contém os genes *A*, *B*, *C* e *D*, então a inserção entre *A* e *D* proporcionaria a ordem *ABCD* ou *DCBA*, dependendo da orientação. Verifique as diferentes orientações das inserções na Figura 5.14.

Como é possível que F se integre em diferentes sítios? Se o DNA de F apresentar uma região homóloga a quaisquer das diversas regiões no cromossomo bacteriano, qualquer uma delas poderia atuar como uma região de pareamento, na qual o pareamento poderia ser seguido por um *crossover*. Atualmente essas regiões de homologia sabidamente são segmentos principalmente de elementos de transposição, denominados *sequências de inserção*. Para uma explicação completa sobre as sequências de inserção, ver Capítulo 15.

Portanto, o fator de fertilidade existe em dois estados:

1. O estado de plasmídio: como um elemento citoplasmático livre, F é facilmente transferido para receptoras F^-.
2. O estado integrado: como uma parte contígua de um cromossomo circular, F é transmitido apenas muito tardiamente na conjugação.

O ciclo de conjugação de *E. coli* está resumido na Figura 5.15.

Mapeamento de cromossomos bacterianos

Mapeamento cromossômico geral por meio da utilização do tempo de entrada.

Wollman e Jacob perceberam que seria fácil a construção de mapas de ligação a partir dos resultados de cruzamento interrompido por meio da utilização, como uma medida de “distância”, dos tempos nos quais os alelos doadores aparecem pela primeira vez após o cruzamento. As unidades de distância de mapa nesse

caso são minutos. Portanto, se b^+ começa a entrar na célula F⁻ 10 minutos após a^+ começar a entrar, a^+ e b^+ estão a uma distância de 10 unidades (ver mapa, anteriormente). Assim como os mapas eucarióticos com base em *crossovers*, esses mapas de ligação eram originalmente construções puramente genéticas. Na ocasião em que foram originalmente planejados, não havia um modo para testar a sua base física.

Mapeamento cromossômico detalhado por meio da utilização da frequência de recombinantes. Para que um exconjugante adquira genes doadores como uma característica permanente do seu genoma, o fragmento doador precisa se recombinar com o cromossomo receptor. Entretanto, observe que o mapeamento do tempo de entrada não tem por base a frequência de recombinantes. De fato, as unidades são minutos, não FR. Não obstante, a frequência de recombinantes pode ser utilizada para um tipo de mapeamento mais meticuloso em bactérias, um método para o qual nos voltamos agora.

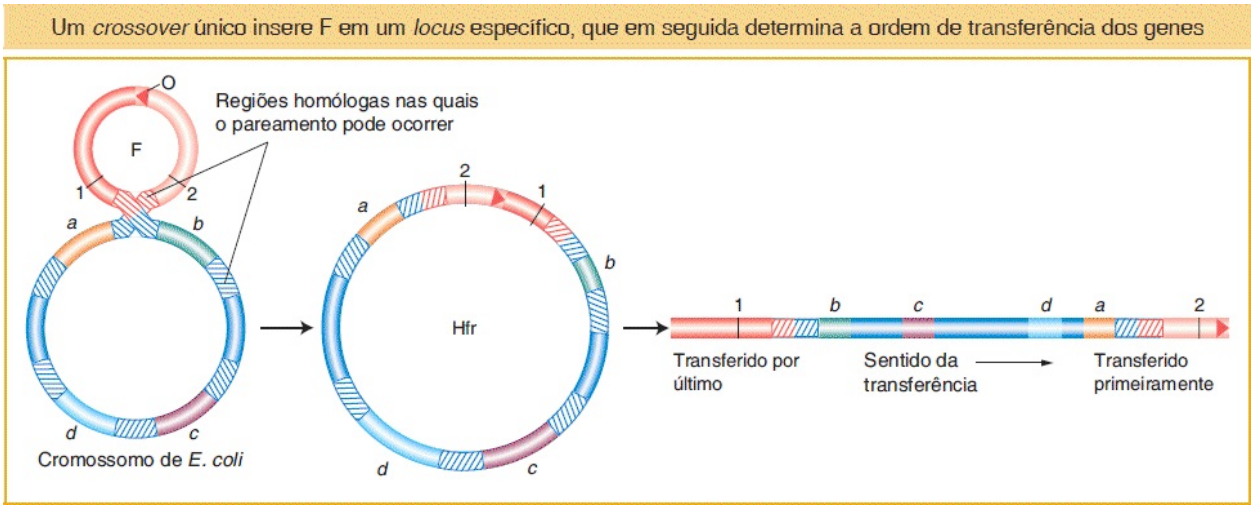


FIGURA 5.13 A inserção de F cria uma célula Hfr. Os marcadores hipotéticos 1 e 2 estão demonstrados em F para ilustrar a direção da inserção. A origem (O) é o ponto de mobilização no qual ocorre a inserção dentro do cromossomo de *E. coli*; a região de pareamento é homóloga a uma região do cromossomo de *E. coli*; *a* a *d* são genes representativos no cromossomo de *E. coli*. As regiões de pareamento (*hachuradas*) são idênticas no plasmídio e no cromossomo. Elas são derivadas de elementos móveis denominados *sequências de inserção* (ver Capítulo 15). Neste exemplo, a célula Hfr criada pela inserção de F transferiria os seus genes na ordem *a*, *d*, *c*, *b*.

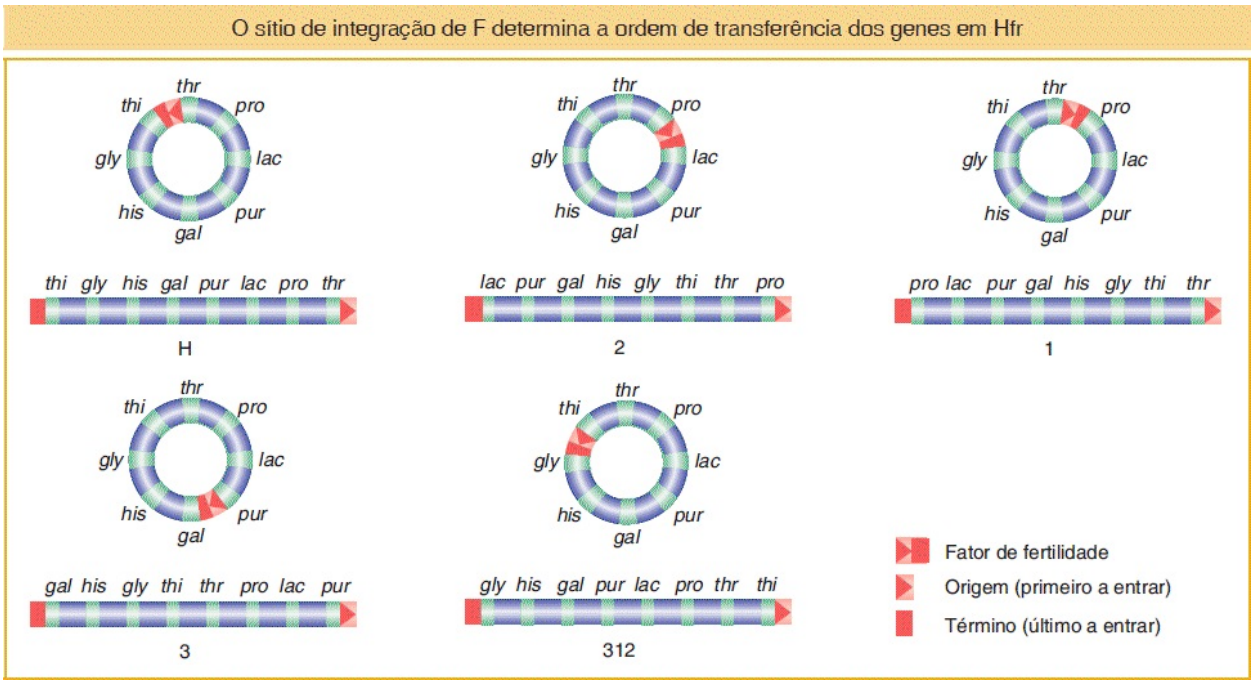


FIGURA 5.14 Cada uma das cinco linhagens Hfr de *E. coli* demonstradas apresentam diferentes pontos de inserção do plasmídio F e orientações. Todas as linhagens apresentam a mesma ordem de genes no cromossomo de *E. coli*. A orientação do fator F determina qual gene entra na célula receptora primeiro. O gene mais próximo do término entra por último.

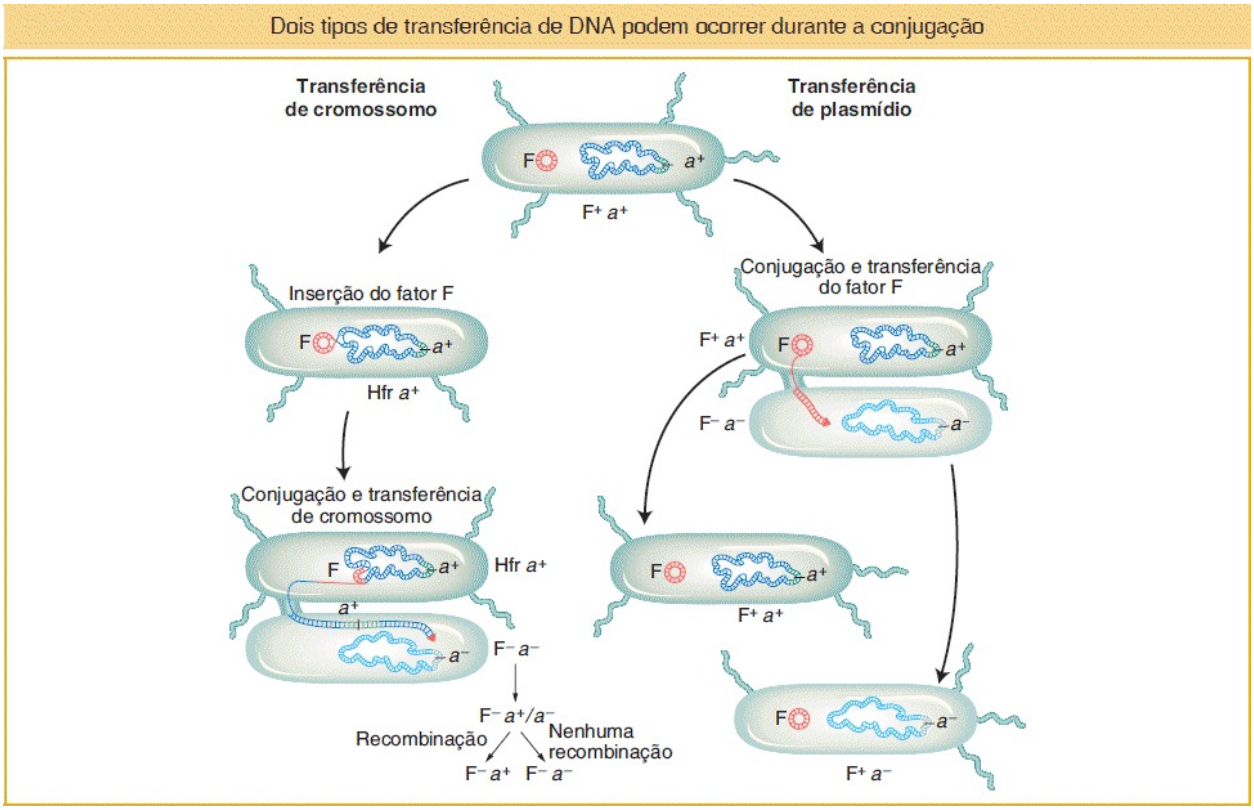


FIGURA 5.15 A conjugação pode ocorrer por meio da transferência parcial de um cromossomo que contém o fator F ou por meio da transferência de um plasmídio F que permanece uma entidade em separado.

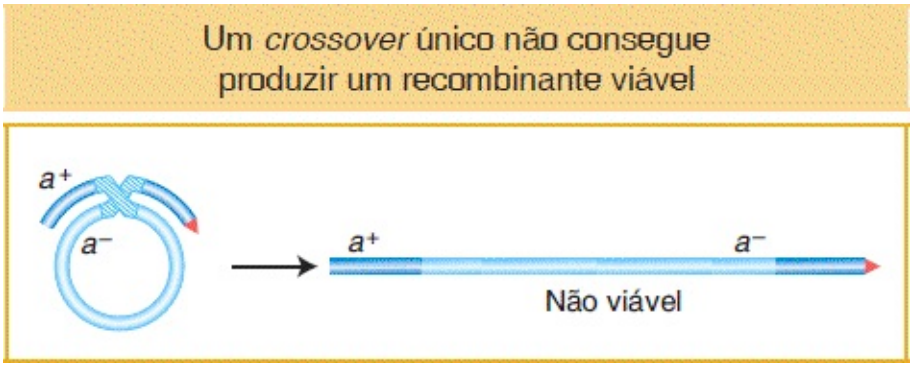
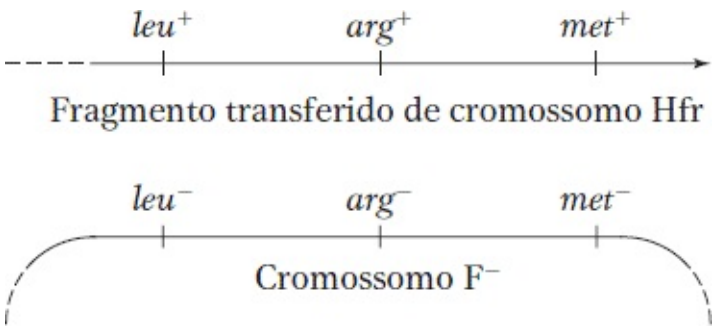


FIGURA 5.16 Um *crossover* único entre o exogenoto e o endogenoto em um merozigoto levaria a um cromossomo linear e parcialmente diploide, que não sobreviveria.

Primeiramente, precisamos compreender algumas características especiais do evento de recombinação em bactérias. Relembre que a recombinação não ocorre entre dois genomas integrais, como ocorre em eucariotos. Contrariamente, ele ocorre entre um genoma *completo*, da célula receptora F^-, denominado **endogenoto**, e um genoma *incompleto*, derivado da célula doadora Hfr e denominado **exogenoto**. A célula nesse estágio apresenta duas cópias de um segmento de DNA: uma cópia é parte do endogenoto e a outra cópia é parte do exogenoto. Portanto, nesse estágio, a célula é uma diploide *parcial*, denominada **merozigoto**. A genética bacteriana é a genética do merozigoto. Um *crossover* único em um merozigoto quebraria o anel e, portanto, não produziria recombinantes viáveis, conforme demonstrado na Figura 5.16. Para manter o círculo intacto, deve haver um número par de *crossovers*. Um número par de *crossovers* produz um cromossomo circular intacto e um fragmento. Embora os referidos eventos de recombinação sejam representados de modo resumido como *crossovers* duplos, o real mecanismo molecular é um pouco diferente, mais como uma invasão do endogenoto por uma seção interna do exogenoto. O outro produto do "*crossover* duplo", o fragmento, em geral é perdido no crescimento celular subsequente. Portanto, apenas um dos produtos recíprocos da recombinação sobrevive. Portanto, outra característica única da recombinação bacteriana é que devemos nos esquecer dos produtos de troca recíproca na maior parte dos casos.

> CONCEITO-CHAVE A recombinação durante a conjugação resulta de um evento semelhante a um *crossover* duplo, que dá origem a recombinantes recíprocos, dos quais apenas um sobrevive.

Com essa compreensão, podemos examinar o mapeamento de recombinação. Suponha que desejamos calcular as distâncias de mapa que separam três *loci* próximos: *met*, *arg* e *leu*. Para examinar a recombinação desses genes, precisamos de "tri-híbridos", exconjugantes que receberam todos os três marcadores doadores. Presuma que um experimento de cruzamento interrompido tenha demonstrado que a ordem é *met*, *arg*, *leu*, com *met* transferido primeiro e *leu* por último. Para obter um tri-híbrido, precisamos do merozigoto diagramado aqui:



Para obter esse merozigoto, primeiramente devemos selecionar exconjugantes estáveis que contenham o *último* alelo doador, o qual, nesse caso, é *leu*$^+$. Por quê? Em exconjugantes *leu*$^+$, sabemos que todos os três marcadores foram transferidos para o receptor, tendo em vista que *leu* é o último alelo doador. Também sabemos que no mínimo o marcador *leu*$^+$ foi integrado ao endogenoto. Desejamos saber com que frequência os outros dois marcadores também foram integrados, de modo que possamos determinar o número de eventos de recombinação nos quais *arg*$^+$ ou *met*$^+$ tenha sido omitido em virtude de *crossover* duplo.

Agora o objetivo é contar as frequências de *crossovers* em locais diferentes. Observe que agora temos uma situação diferente da análise de conjugação

interrompida. No mapeamento por meio de conjugação interrompida, medimos o tempo de entrada de *loci* individuais; para ser herdado de modo estável, cada marcador tem de se recombinar com o cromossomo do receptor por meio de um *crossover* duplo que o abranja. Entretanto, na análise da frequência de recombinantes, selecionamos especificamente tri-híbridos como um ponto de partida e agora precisamos considerar as diversas combinações possíveis dos três alelos doadores que podem ser inseridos por meio de *crossover* duplo nos diversos intervalos. Sabemos que leu^+ necessariamente penetrou e se inseriu pois o selecionamos, mas os recombinantes leu^+ que selecionamos podem ou não ter incorporado os outros marcadores doadores, dependendo de onde ocorreu o *crossover* duplo. Portanto, o procedimento é primeiramente selecionar exconjugantes leu^+ e em seguida isolar e testar uma amostra grande deles, para verificar quais dos outros marcadores foram integrados.

Vejamos um exemplo. No cruzamento Hfr $leu^+\ arg^+\ met^+\ str^s$ × F$^-$ $met^-\ arg^-\ leu^-\ str^r$, selecionaríamos os recombinantes leu^+ e em seguida os examinaríamos em relação aos alelos arg^+ e met^+, denominados **marcadores não selecionados**. A Figura 5.17 ilustra os tipos de eventos de *crossover* duplo esperados. Um *crossover* necessariamente ocorreu no lado esquerdo do marcador *leu* e o outro obrigatoriamente ocorreu no lado direito. Presumiremos que os exconjugantes leu^+ sejam dos tipos e das frequências a seguir:

$leu^+\ arg^-\ met^-$ 4%
$leu^+\ arg^+\ met^-$ 9%
$leu^+\ arg^+\ met^+$ 87%

Os *crossovers* duplos necessários para produzir esses genótipos estão demonstrados na Figura 5.17. Os primeiros dois tipos são a chave, tendo em vista que necessitam de um *crossover* entre *leu* e *arg* no primeiro caso e entre *arg* e *met* no segundo. Portanto, as frequências relativas desses tipos correspondem aos tamanhos dessas duas regiões entre os genes. Concluiríamos que a região de *leu* a *arg* tem 4 u.m. e que a região de *arg* a *met* tem 9. u.m.

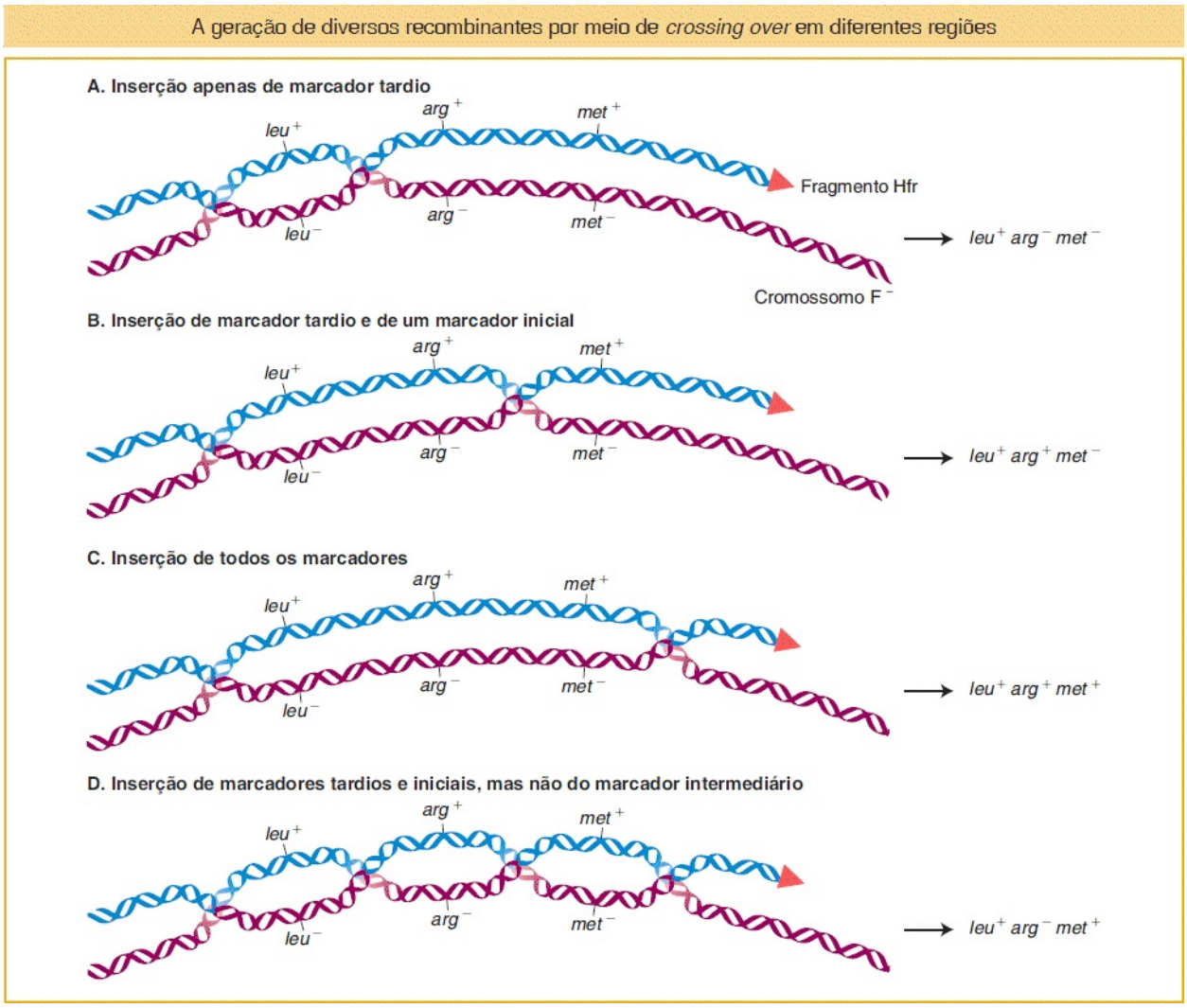


FIGURA 5.17 O diagrama demonstra como genes podem ser mapeados por meio de recombinação em *E. coli*. Em exconjugantes, a seleção é realizada em relação a merozigotos que contenham o marcador *leu*$^+$, que é doado tardiamente. Os marcadores iniciais (*arg*$^+$ e *met*$^+$) podem ou não ser inseridos, dependendo do sítio onde ocorre a recombinação entre o fragmento Hfr e o cromossomo F$^-$. As frequências dos eventos diagramados nas partes *a* e *b* são utilizadas para obter os tamanhos relativos das regiões de *leu* a *arg* e *arg* a *met*. Observe que, em cada caso, apenas o DNA inserido no cromossomo F$^-$ sobrevive; o outro fragmento é perdido.

Em um cruzamento tal como o descrito há pouco, um tipo de possíveis recombinantes de genótipo *leu*$^+$ *arg*$^-$ *met*$^+$ requer quatro *crossovers* em vez de dois (ver parte inferior da Figura 5.17). Esses recombinantes raramente são recuperados, tendo em vista que a sua frequência é muito baixa em comparação àquela dos outros tipos de recombinantes.

Plasmídios F carreadores de fragmentos genômicos

O fator F em linhagens Hfr em geral é consideravelmente estável em sua posição de inserção. Entretanto, ocasionalmente um fator F claramente sai do cromossomo graças à reversão do processo de recombinação que o inseriu primeiramente. As duas regiões homólogas de pareamento em ambos os lados realizam um novo pareamento e ocorre um *crossover* para liberar o plasmídio F. Entretanto, algumas vezes a saída não é limpa e o plasmídio carreia com ele uma parte do cromossomo bacteriano. Um plasmídio F que carreia um DNA genômico bacteriano é denominado **plasmídio F' (*prime* F)**.

A primeira evidência desse processo teve origem nos experimentos em 1959 por Edward Adelberg e François Jacob. Uma de suas observações-chave foi de uma Hfr na qual o fator F estava integrado próximo do *locus* lac^+. Iniciando com essa linhagem Hfr lac^+, Jacob e Adelberg observaram um derivado F^+ que, em cruzamentos, transferia lac^+ para receptores $F^- lac^-$ a uma frequência muito alta. (Essas transferentes puderam ser detectados por plaqueamento em meio contendo lactose, no qual apenas lac^+ pode crescer.) O lac^+ transferido não é incorporado ao cromossomo principal do receptor, o qual sabemos que retém o alelo lac^-, tendo em vista que esses exconjugantes F^+ lac^+ ocasionalmente dão origem a células-filhas $F^- lac^-$, a uma frequência de 1×10^{-3}. Portanto, o genótipo desses receptores aparentava ser F' lac^+/F^- lac^-. Em outras palavras, os exconjugantes lac^+ aparentavam carrear um plasmídio F' com uma parte do cromossomo doador incorporado. A origem desse plasmídio F' está demonstrada na Figura 5.18. Observe que a excisão defeituosa ocorre em virtude da existência de outra região homóloga próxima, que pareia com a original. O F' em nosso exemplo é denominado F' *lac*, tendo em vista que a parte do cromossomo hospedeiro que é captada possui o gene *lac*. Observou-se que os fatores F' carreiam muitos genes cromossômicos diferentes e foram denominados de acordo. Por exemplo, os fatores F' que carreiam *gal* ou *trp* são denominados F' *gal* e F' *trp*, respectivamente. Tendo em vista que as células F lac^+/F^- lac^- são de fenótipo lac^+, sabemos que lac^+ é dominante em relação a lac^-.

Diploides parciais produzidos com a utilização de linhagens F' são úteis para

alguns aspectos da genética bacteriana de rotina, tais como o estudo da dominância ou da interação de alelos. Algumas linhagens F' conseguem carrear partes muito grandes (até 25%) do cromossomo bacteriano.

> CONCEITO-CHAVE O DNA de um plasmídio F' contém parte do fator F e parte do genoma bacteriano. Assim como os plasmídios F, os plasmídios F' são transferidos rapidamente. Eles podem ser utilizados para estabelecer diploides parciais para estudos de dominância bacteriana e interação de alelos.

Alça defeituosa produz F', um plasmídio F que contém DNA cromossômico

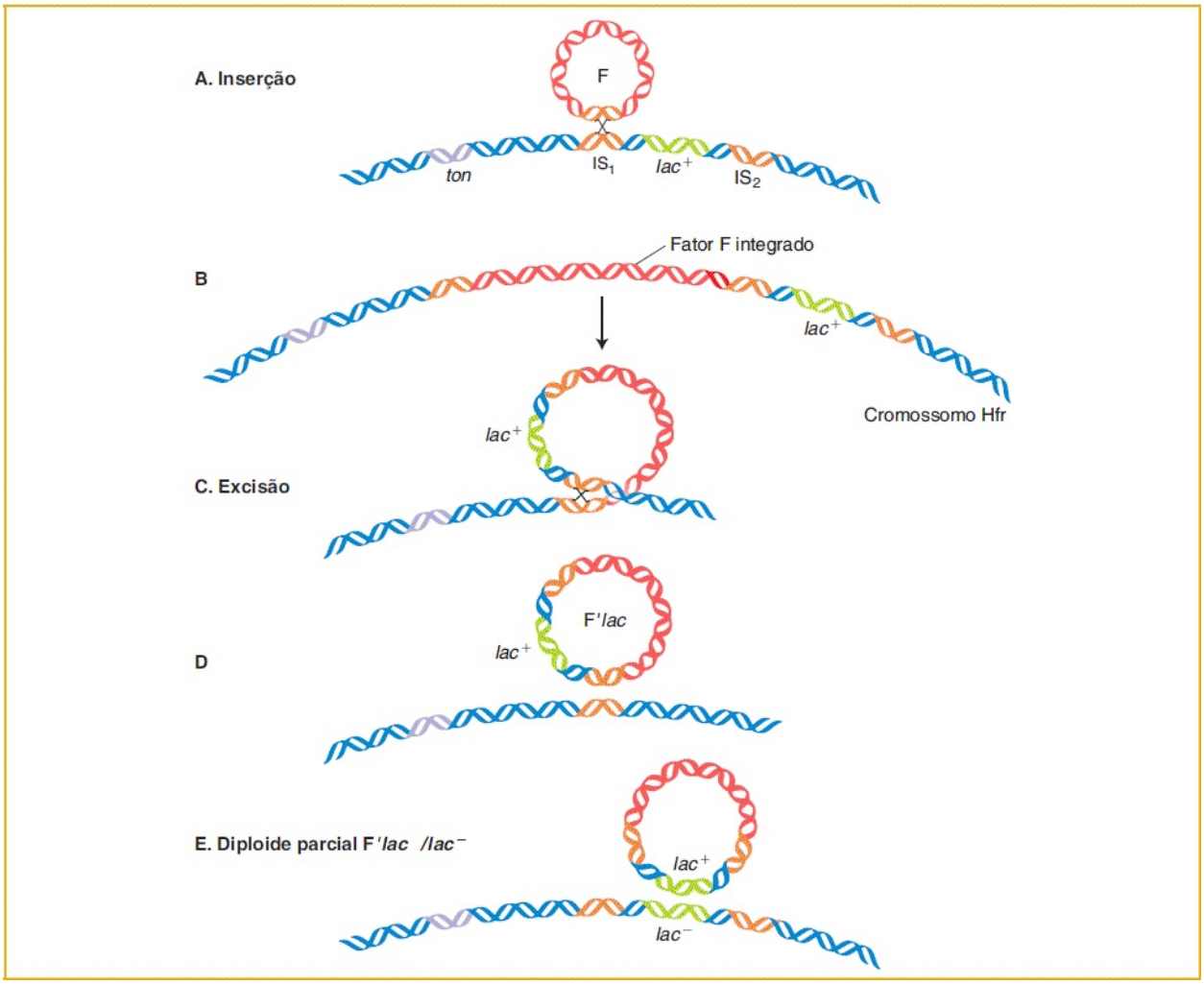


FIGURA 5.18 Um fator F consegue captar o DNA cromossômico quando ele sai de um cromossomo. **A.** O F é inserido em uma linhagem Hfr em um elemento repetitivo, identificado como IS_1 (sequência de inserção 1), entre os alelos *ton* e lac^+. **B.** O fator F inserido. **C.** "Alça" anormal por *crossing over* com um elemento diferente, IS_2, para incluir o *locus lac*. **D.** A partícula F' lac^+ resultante. **E.** Diploide parcial F' lac^+/F^- lac^- produzido por meio da transferência da partícula F' lac^+ para uma receptora F^- lac^-. (Dados de G. S. Stent and R. Calendar, *Molecular Genetics*, 2nd ed.)

Plasmídios R

Uma propriedade alarmante das bactérias patogênicas foi verificada pela primeira vez por meio de estudos em hospitais japoneses na década de 1950. A disenteria bacteriana é causada por bactérias do gênero *Shigella*. Essa bactéria inicialmente era sensível a uma ampla variedade de antibióticos que eram utilizados para controlar a doença. Entretanto, nos hospitais japoneses, *Shigella* isoladas de pacientes com disenteria eram simultaneamente resistentes a muitos desses fármacos, incluindo penicilina, tetraciclina, sulfanilamida, estreptomicina e cloranfenicol. Essa resistência a múltiplos fármacos foi herdada como um pacote genético único e pode ser transmitida de modo infeccioso — não apenas para outras linhagens de *Shigella* sensíveis, mas também para outras espécies de bactérias correlatas. Esse talento, que se assemelha à mobilidade do plasmídio F de *E. coli*, é um talento extraordinariamente útil para a bactéria patogênica, tendo em vista que a resistência pode ser difundida rapidamente em uma população. Entretanto, suas implicações para a ciência médica são terríveis, tendo em vista que a doença bacteriana subitamente se torna resistente ao tratamento com uma grande variedade de fármacos.

Do ponto de vista do geneticista, entretanto, o mecanismo comprovou ser interessante e útil na engenharia genética. Os vetores que carreiam essas resistências múltiplas comprovaram ser outro grupo de plasmídios, denominados **plasmídios R**. Eles são transferidos rapidamente na conjugação celular, de modo muito semelhante ao plasmídio F em *E. coli*.

De fato, os plasmídios R em *Shigella* comprovaram ser justamente os primeiros de muitos elementos genéticos semelhantes a serem descobertos. Todos existem no estado de plasmídio no citoplasma. Observou-se que esses elementos carreiam muitos tipos diferentes de genes em bactérias. A Tabela 5.2 demonstra

algumas das características que podem ser transmitidas pelos plasmídios. A Figura 5.19 demonstra um exemplo de um plasmídio bem-conhecido, isolado da indústria de laticínios.

Tabela 5.2 Determinantes genéticos transmitidos por plasmídios.

Característica	Exemplos de plasmídios
Fertilidade	F, R1, Col
Produção de bacteriocina	Col E1
Resistência a metais pesados	R6
Produção de enterotoxina	Ent
Metabolismo da cânfora	Cam
Tumorigenicidade em plantas	T1 (em *Agrobacterium tumefaciens*)

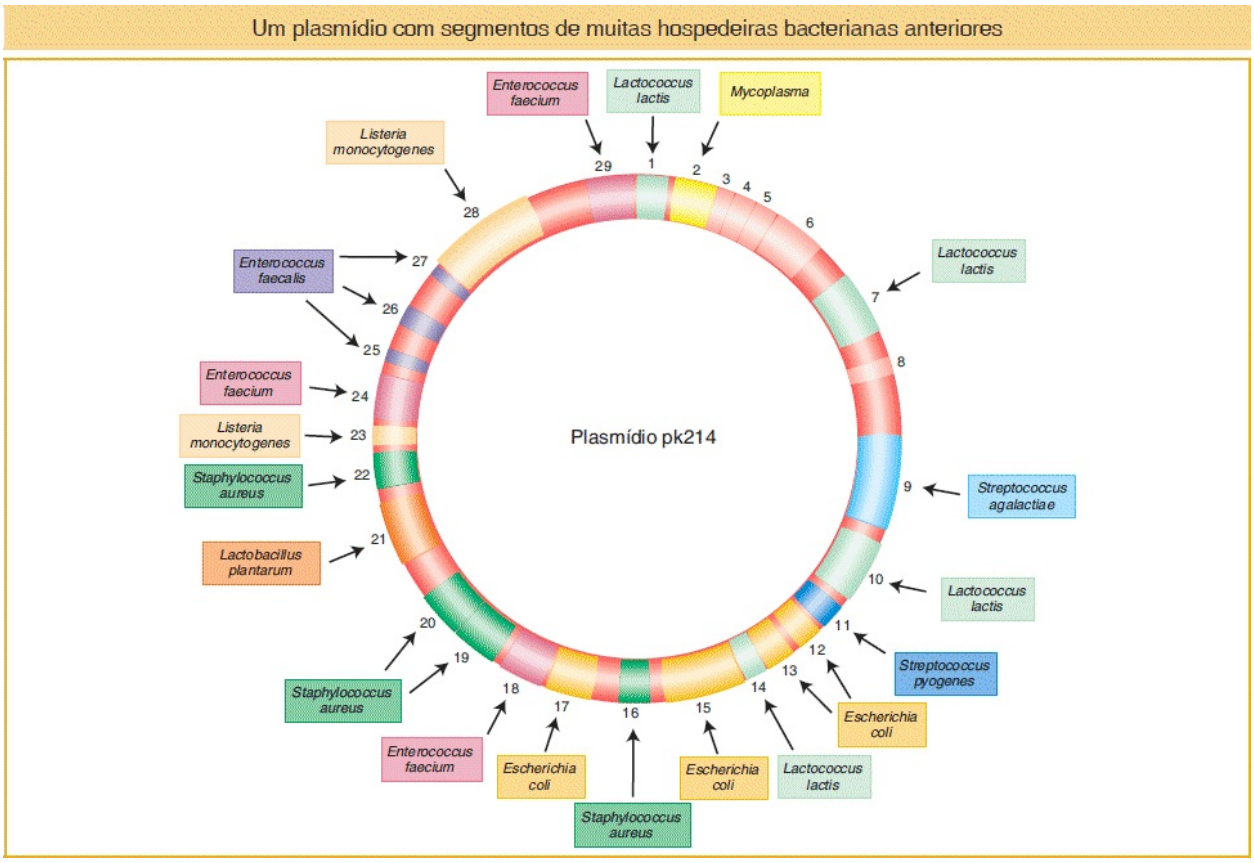


FIGURA 5.19 O diagrama demonstra as origens dos genes do plasmídio pk214 de *Lactococcus lactis*. Os genes são provenientes de muitas bactérias diferentes. (Dados da Tabela 1 em V. Perreten, F. Schwartz, L. Cresta, M. Boeglin, G. Dasen and M. Teuber, *Nature* 389, 1997, 801-802.)

Derivados modificados de plasmídios R, tais como pBR 322 e pUC (ver Capítulo 10), tornam-se os vetores preferidos para a clonagem molecular do DNA de todos os organismos. Os genes em um plasmídio R que conferem resistência podem ser utilizados como marcadores para o seguimento da movimentação dos vetores entre as células.

Nos plasmídios R, os alelos para resistência a antibióticos com frequência estão contidos em uma unidade denominada *transpóson* (Figura 5.20). Transpósons são segmentos únicos de DNA que conseguem mover-se para sítios diferentes no genoma, um processo denominado transposição. (Os mecanismos de transposição, a qual ocorre na maior parte das espécies estudadas, serão detalhados no Capítulo 15.) Quando um transpóson no genoma se movimenta até um local novo, ocasionalmente ele consegue englobar entre as suas extremidades

diversos tipos de genes, incluindo alelos para resistência a fármacos, e os transporta até seus novos locais como passageiros. Por vezes, um transpóson carreia um alelo de resistência a fármacos até um plasmídio, criando um plasmídio R. Assim como os F, muitos plasmídios R são conjugativos; em outras palavras, são efetivamente transmitidos para uma célula receptora durante a conjugação. Até mesmo os plasmídios R que não são conjugativos e que nunca deixam suas próprias células podem doar seus alelos R para um plasmídio conjugativo por transposição. Portanto, por meio dos plasmídios, os alelos de resistência a antibióticos podem ser difundidos rapidamente em uma população de bactérias. Embora a difusão de plasmídios R seja uma estratégia eficaz para a sobrevivência das bactérias, ela apresenta um problema importante para a prática médica, conforme mencionado anteriormente, tendo em vista que as populações bacterianas rapidamente se tornam resistentes a qualquer novo fármaco antibiótico que seja inventado e aplicado em humanos.

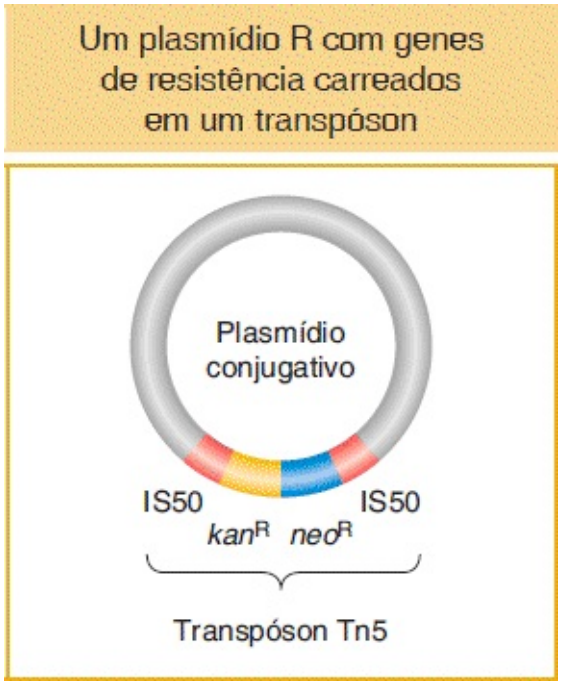


FIGURA 5.20 Um transpóson, tal como Tn5, pode adquirir diversos genes de resistência a fármacos (neste caso, aqueles em relação à resistência aos fármacos canamicina e neomicina) e os transmitir rapidamente para um plasmídio, levando à transferência infecciosa de genes de resistência como um pacote. A sequência de inserção 50 (IS50) forma os lados de TN5.

5.3 Transformação bacteriana

Algumas bactérias conseguem captar fragmentos de DNA do meio externo e a referida captação constitui outro modo por meio do qual as bactérias conseguem trocar seus genes. A fonte de DNA pode ser outras células da mesma espécie ou células de outras espécies. Em alguns casos, o DNA foi liberado a partir de células mortas; em outros casos, o DNA foi secretado de células bacterianas vivas. O DNA captado integra-se ao cromossomo da receptora. Se esse DNA for de um genótipo diferente daquele da receptora, o genótipo da receptora pode ser permanentemente alterado, um processo adequadamente denominado **transformação**.

Natureza da transformação

A transformação foi descoberta na bactéria *Streptococcus pneumoniae* em 1928 por Frederick Griffith. Posteriormente, em 1944, Oswald T. Avery, Colin M. MacLeod e Maclyn McCarty demonstraram que o "princípio transformante" era o DNA. Ambos os resultados são marcos na elucidação da natureza molecular dos genes. Consideramos esse trabalho em mais detalhes no Capítulo 7.

O DNA transformante é incorporado ao cromossomo bacteriano por meio de um processo análogo aos eventos de recombinação dupla observados em cruzamentos Hfr × F⁻. Entretanto, observe que, na *conjugação*, o DNA é transferido de uma célula viva para outra por meio do contato próximo, enquanto na *transformação*, fragmentos isolados de DNA externo são captados por uma célula por meio da parede celular e da membrana plasmática. A Figura 5.21 demonstra um modo por meio do qual esse processo pode ocorrer.

A transformação tem sido útil em diversas áreas da pesquisa bacteriana, tendo em vista que o genótipo de uma linhagem pode ser deliberadamente alterado em um modo muito específico por meio da transformação com um fragmento de DNA apropriado. Por exemplo, a transformação é amplamente utilizada na engenharia genética. Observou-se que até mesmo células eucarióticas podem ser transformadas, por meio da utilização de procedimentos consideravelmente semelhantes, e essa técnica tem sido inestimável para a modificação de células

eucarióticas (ver Capítulo 10).

Mapeamento cromossômico com a utilização de transformação

A transformação pode ser utilizada para medir quão próximos dois genes estão ligados em um cromossomo bacteriano. Quando o DNA (o cromossomo bacteriano) é extraído para experimentos de transformação, alguma quebra em fragmentos menores é inevitável. Se dois genes do doador estiverem localizados muito próximos no cromossomo, existe uma boa chance de que por vezes eles venham a ser carreados no mesmo fragmento de DNA transformante. Portanto, ambos serão captados, causando uma **transformação dupla**. Contrariamente, se os genes estiverem amplamente separados no cromossomo, mais provavelmente eles serão carreados em segmentos transformantes separados. Um genoma poderia captar ambos os segmentos de modo independente, criando um transformante duplo, mas aquele desfecho não é provável. Portanto, em genes amplamente separados, a frequência de transformantes duplos será igual ao produto das frequências dos transformantes únicos. Portanto, deve ser possível testar a ligação próxima testando o desvio pela regra do produto. Em outras palavras, se os genes estiverem ligados, a proporção de transformantes duplos será superior ao produto das frequências de transformantes únicos.

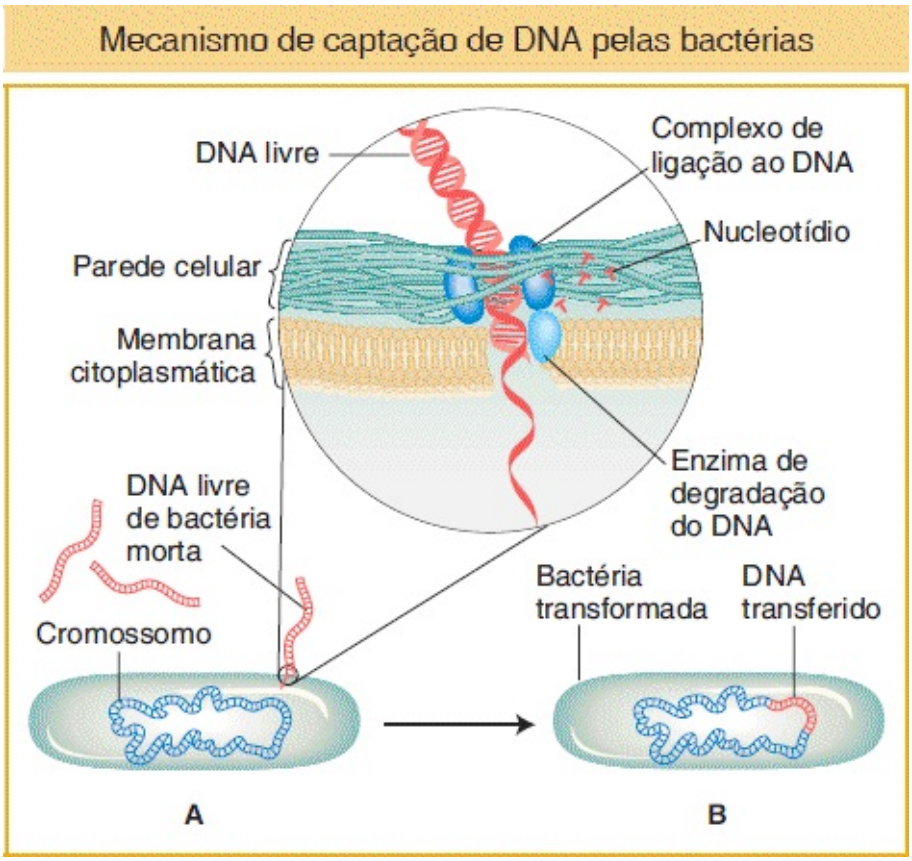


FIGURA 5.21 Uma bactéria sofrendo transformação (**A**) capta DNA livre liberado de uma célula bacteriana morta. Na medida em que os complexos de ligação ao DNA na superfície bacteriana captam o DNA (*inserção*), enzimas degradam um filamento em nucleotídios; um derivado do outro filamento pode integrar-se ao cromossomo bacteriano (**B**).

Infelizmente, a situação se torna mais complexa por diversos fatores — o mais importante sendo que nem todas as células em uma população de bactérias são competentes para serem transformadas. Não obstante, no final deste capítulo, você poderá aguçar as suas habilidades na análise de transformação em um dos problemas, o qual presume que 100% das células receptoras sejam competentes.

> CONCEITO-CHAVE As bactérias conseguem captar fragmentos de DNA a partir do meio adjacente. Dentro da célula, esses fragmentos conseguem integrar-se ao cromossomo.

5.4 Genética de bacteriófagos

A palavra *bacteriófago*, que é uma denominação dos vírus bacterianos, significa "comedor de bactérias". Esses vírus parasitam e matam as bactérias. Trabalhos pioneiros sobre a genética dos bacteriófagos em meados do século 20 formaram o fundamento de pesquisas mais recentes sobre vírus que causam tumores e outros tipos de vírus de animais e plantas. Desse modo, os vírus bacterianos proporcionaram um importante sistema-modelo.

Esses vírus podem ser utilizados em dois tipos diferentes de análises genéticas. Primeiramente, dois genótipos de fagos distintos podem ser cruzados para medir a recombinação e, portanto, mapear o genoma viral. O mapeamento do genoma viral por meio desse método é o tópico desta seção. Em segundo lugar, os bacteriófagos podem ser utilizados como um modo de unir genes bacterianos para estudos de ligação e outros estudos genéticos. Estudaremos a utilização dos fagos em estudos bacterianos na Seção 5.5. Além disso, conforme veremos no Capítulo 10, os fagos são utilizados em tecnologia de DNA como portadores, ou vetores, de DNA estranho. Antes que possamos compreender a genética dos fagos, devemos primeiramente examinar o ciclo de infecção dos fagos.

Infecção de bactérias por fagos

A maior parte das bactérias é suscetível ao ataque por bacteriófagos. Um fago é composto por um "cromossomo" de ácido nucleico (DNA ou RNA), circundado por um revestimento de moléculas proteicas. Os tipos de fagos são identificados não pelos nomes das espécies, mas por símbolos — por exemplo, fago T4, fago λ e assim por diante. As Figuras 5.22 e 5.23 demonstram a estrutura do fago T4. Durante a infecção, um fago adere a uma bactéria e injeta o seu material genético no citoplasma bacteriano, conforme diagramado na Figura 5.22. Uma micrografia eletrônica do processo está demonstrada na Figura 5.24. As informações genéticas do fago em seguida assumem o controle do maquinário da célula bacteriana ao desligar a síntese dos componentes bacterianos e ao redirecionar o maquinário de síntese bacteriano para produzir os componentes do fago. As cabeças de fagos recém-formadas são dotadas individualmente com réplicas do

cromossomo do fago. Finalmente, muitos descendentes do fago são produzidos e liberados quando a parede celular bacteriana se rompe. Esse processo de ruptura é denominado **lise**. A população de progênie de fagos é denominada **lisado** de fagos.

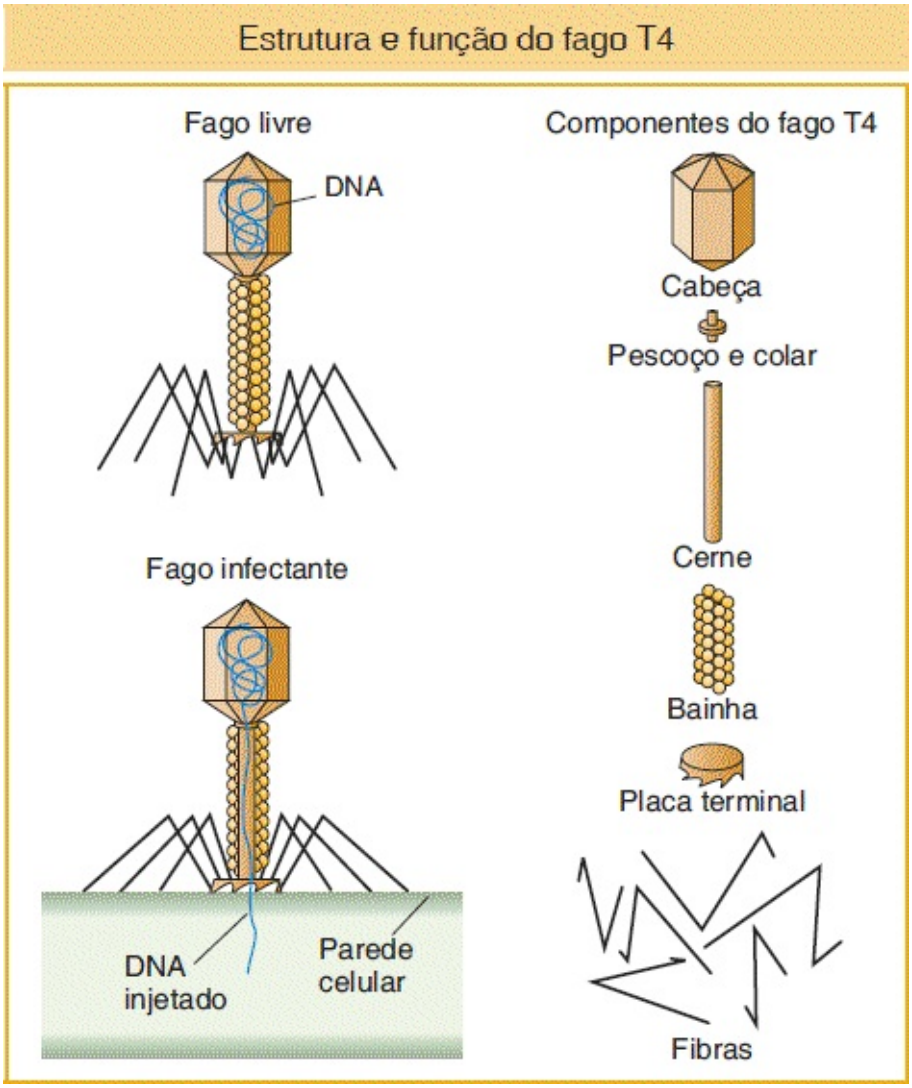


FIGURA 5.22 Um fago infectante injeta DNA por meio de sua estrutura central na célula (*esquerda*). O bacteriófago T4 é demonstrado como um fago livre e em seguida no processo de infecção de uma célula de *E. coli* (*direita*). Os principais componentes estruturais de T4.

Micrografia eletrônica do fago T4

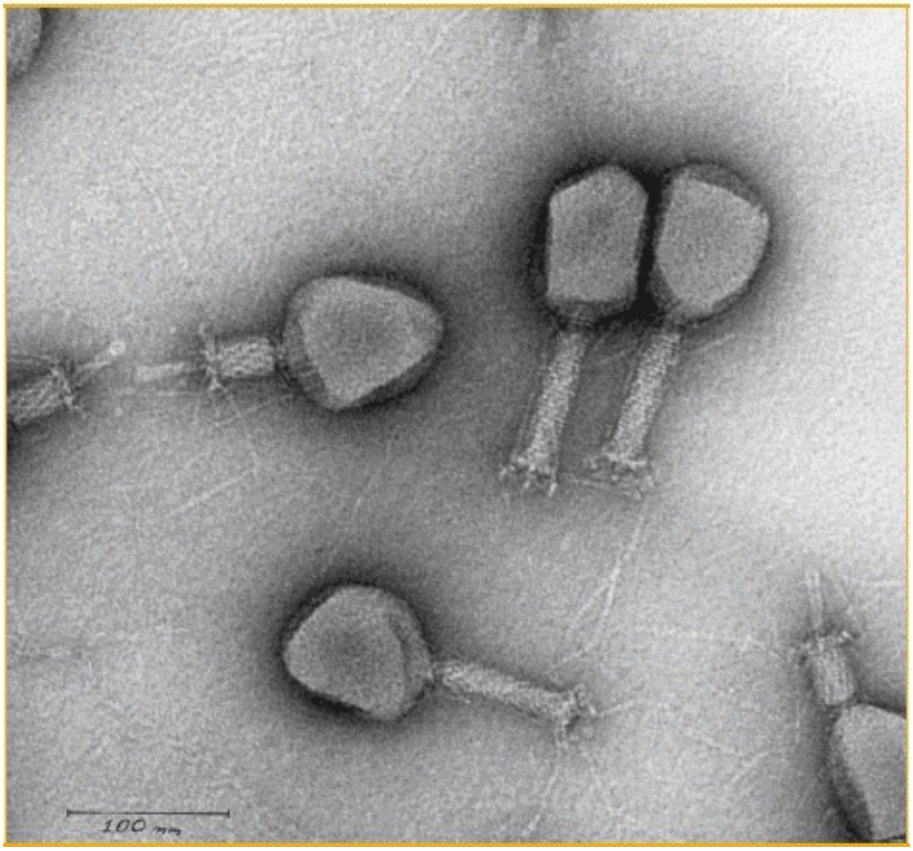


FIGURA 5.23 A ampliação do fago T4 de *E. coli* revela detalhes da cabeça, da cauda e das fibras da cauda. (Science Source.)

Micrografia eletrônica da infecção por fagos

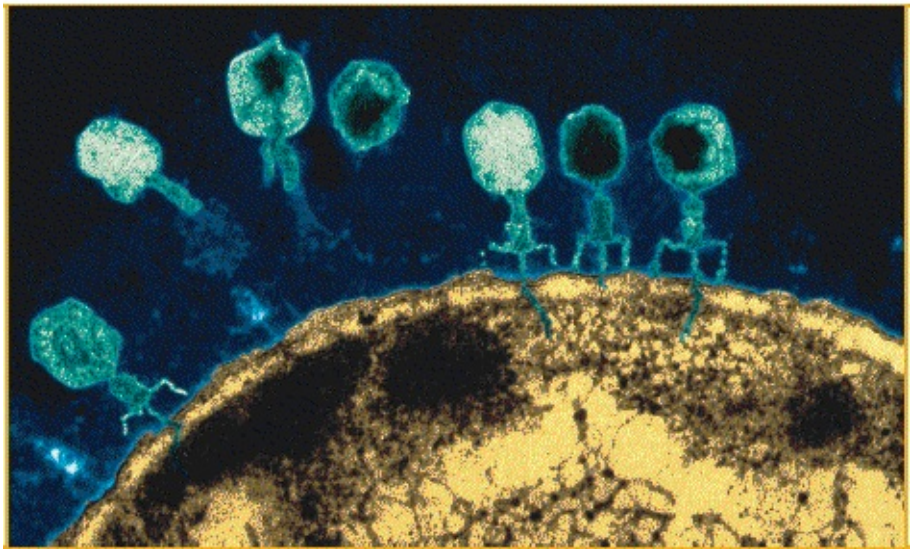


FIGURA 5.24 Bacteriófagos são demonstrados em diversos estágios do processo de infecção, que inclui a ligação e a injeção do DNA. (©Eye of Science/Science Source.)

Como podemos estudar a herança nos fagos quando eles são tão pequenos que são visíveis apenas ao microscópio eletrônico? Nesse caso, não conseguimos produzir uma colônia visível por meio de plaqueamento, mas podemos produzir uma manifestação visível de um fago ao tirar vantagem de diversas características do fago.

Vejamos as consequências da infecção de uma célula bacteriana única por um fago. A Figura 5.25 demonstra a sequência de eventos no ciclo infeccioso que leva à liberação da progênie de fagos pela célula lisada. Após a lise, os fagos da progênie infectam as bactérias vizinhas. Esse ciclo é repetido durante rodadas progressivas de infecção e na medida em que esses ciclos se repetem, a quantidade de células lisadas aumenta exponencialmente. Dentro de 15 horas após a infecção de uma única célula bacteriana por uma única partícula de fago, os efeitos são visíveis a olho nu como uma área clara, ou **placa**, na camada opaca de bactérias que recobre a superfície de uma placa de meio sólido (Figura 5.26). As referidas placas podem ser grandes ou pequenas, difusas ou bem-definidas e assim por diante, dependendo do genótipo do fago. Portanto, a *morfologia da placa* é uma característica do fago, que pode ser analisada no nível genético. Outro fenótipo do fago que podemos analisar geneticamente é a *variedade de hospedeiros*, tendo em vista que os fagos diferem nos espectros de linhagens bacterianas que conseguem infectar e lisar. Por exemplo, uma linhagem específica de bactérias pode ser imune ao fago 1, mas suscetível ao fago 2.

Ciclo do fago que lisa as células hospedeiras

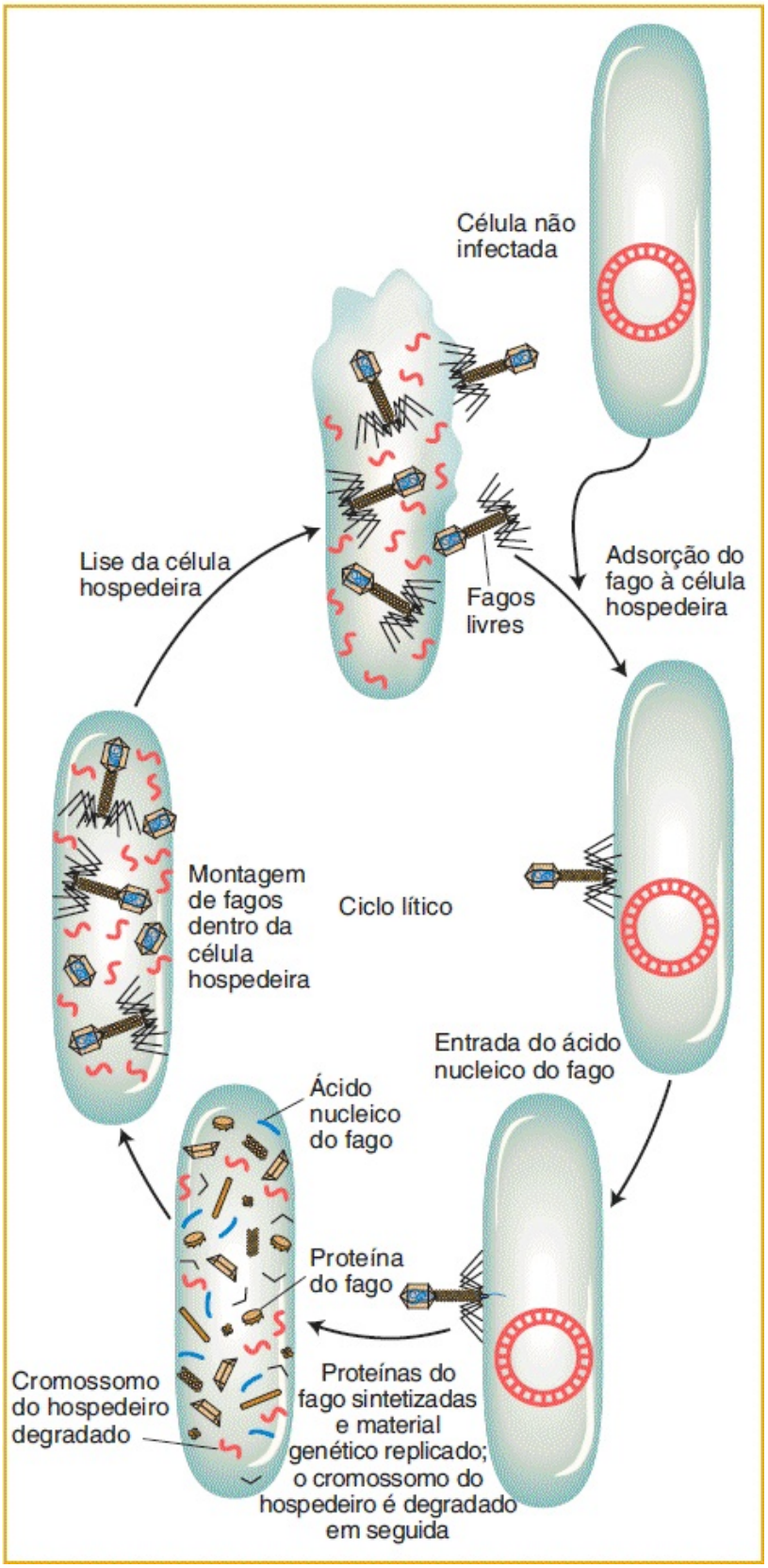


FIGURA 5.25 A infecção por um fago único redireciona o maquinário da célula para a produção da progênie de fagos, que são liberados na lise.

Mapeamento de cromossomos de fagos por meio da utilização de cruzamentos de fagos

Dois genótipos de fagos podem ser cruzados em grande parte do mesmo modo que cruzamos organismos. Um cruzamento de fagos pode ser ilustrado por meio de um cruzamento de fagos T2 originalmente estudados por Alfred Hershey. Os genótipos das duas linhagens parentais no cruzamento de Hershey foram $h^{-}\ r^{+} \times h^{+}\ r^{-}$. Os alelos correspondem aos fenótipos a seguir:

h^{-} : consegue infectar duas linhagens diferentes de *E. coli* (as quais podemos denominar linhagens 1 e 2)

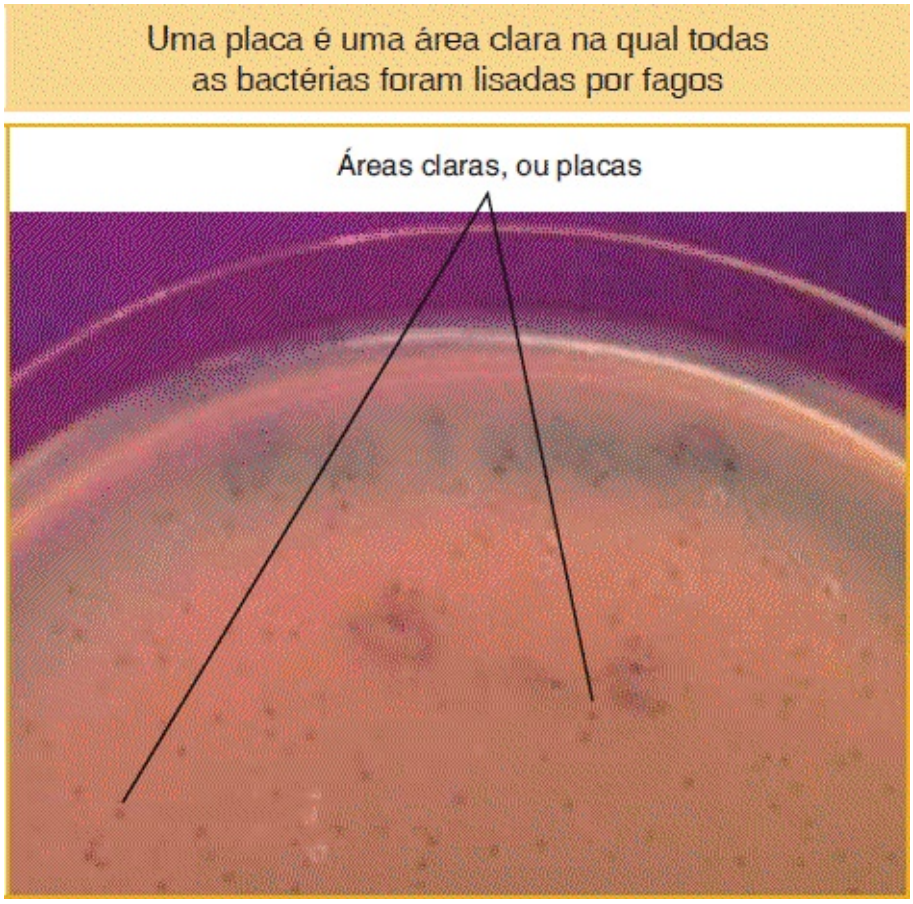


FIGURA 5.26 Por meio de infecção repetida e produção de progênie de fagos, um único fago produz uma

área clara, ou placa, na camada opaca de células bacterianas. (D. Sue Katz, Rogers State University, Claremore, OK.)

h^+ : consegue infectar apenas a linhagem 1
r^- : lisa as células rapidamente, produzindo, assim, placas grandes
r^+: lisa as células lentamente, produzindo placas pequenas

Para realizar o cruzamento a linhagem 1 de *E. coli* é infectada com ambos os genótipos parentais do fago T2. Esse tipo de infecção é denominado **infecção mista** ou **infecção dupla** (Figura 5.27). Após um período de incubação apropriado, o lisado de fagos (que contém os fagos da progênie) é analisado por meio da sua difusão sobre uma camada de bactérias composta por uma mistura das linhagens 1 e 2 de *E. coli*. Quatro tipos de placas são distinguíveis em seguida (Figura 5.28). As placas grandes indicam lise rápida (r^-) e as placas pequenas indicam lise lenta (r^+). As placas de fagos com o alelo h^- infectarão ambas os hospedeiros, formando uma placa clara, enquanto as placas de fagos com o alelo h^+ infectarão apenas um hospedeiro, formando uma placa turva. Portanto, os quatro genótipos podem ser facilmente classificados como parentais (h^- r^+ e h^+ r^-) e recombinantes (h^+ r^+ *e* h^- r^-) e uma frequência de recombinantes pode ser calculada como segue:

$$\text{FR} = \frac{(h^+ r^+) + (h^- r^-)}{\text{placas totais}}$$

Se presumirmos que os cromossomos recombinantes de fagos são lineares, então *crossovers* únicos produzem produtos recíprocos viáveis. Entretanto, os cruzamentos de fagos estão sujeitos a algumas complicações analíticas. Primeiramente, diversas rodadas de troca podem ocorrer no hospedeiro: um recombinante produzido logo após a infecção pode sofrer recombinação adicional na mesma célula ou em ciclos de infecção posteriores. Em segundo lugar, a recombinação pode ocorrer entre fagos geneticamente semelhantes, bem como entre tipos diferentes. Portanto, se considerarmos P_1 e P_2 uma referência aos genótipos parentais gerais, podem ocorrer cruzamentos de $P_1 \times P_1$ e $P_2 \times P_2$, além de $P_1 \times P_2$. Em virtude de ambos esses motivos, os recombinantes de cruzamentos

de fagos são uma consequência de uma *população* de eventos, em vez de eventos definidos de troca de etapa única. Não obstante, *se restante for constante*, o cálculo da FR representa um índice válido da distância de mapa nos fagos.

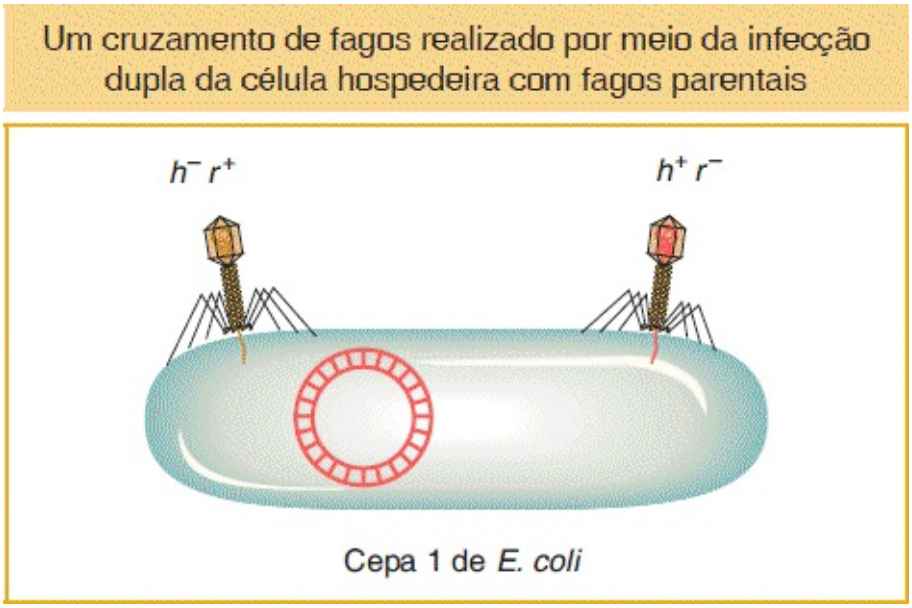


FIGURA 5.27.

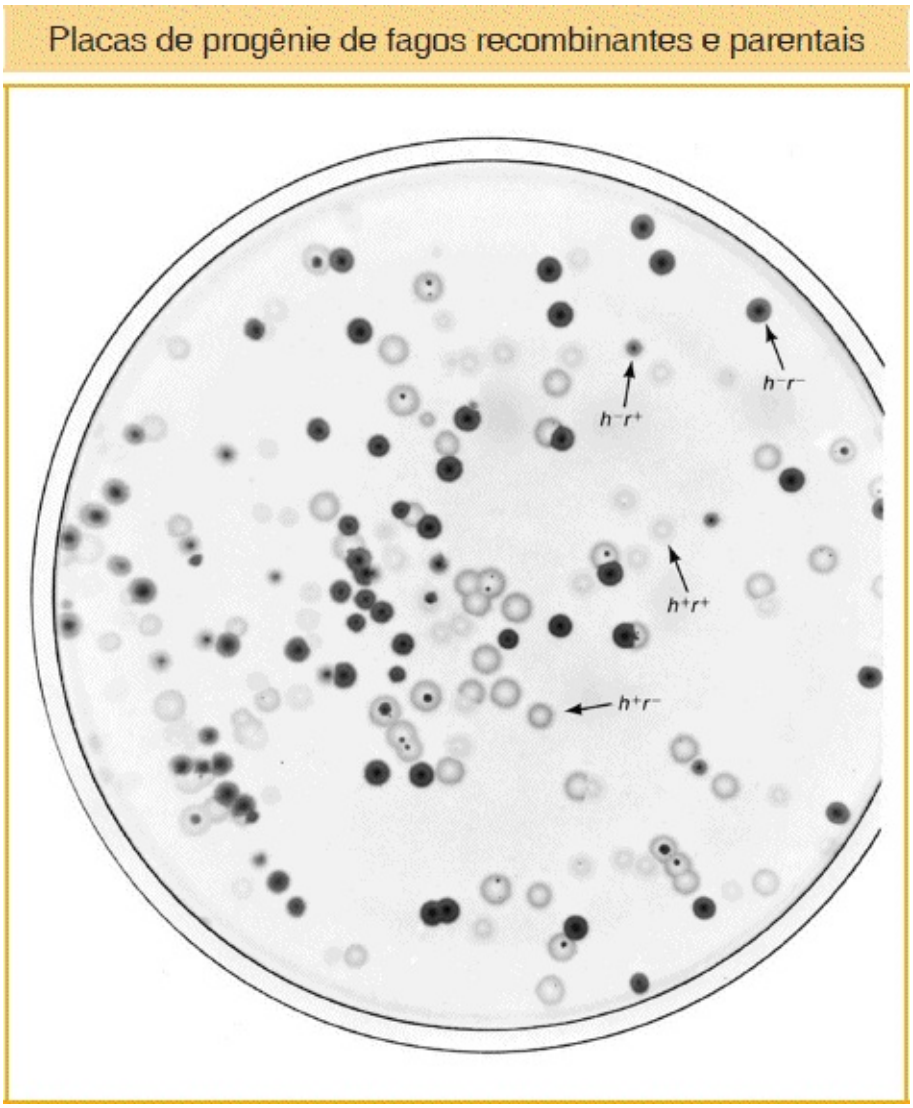
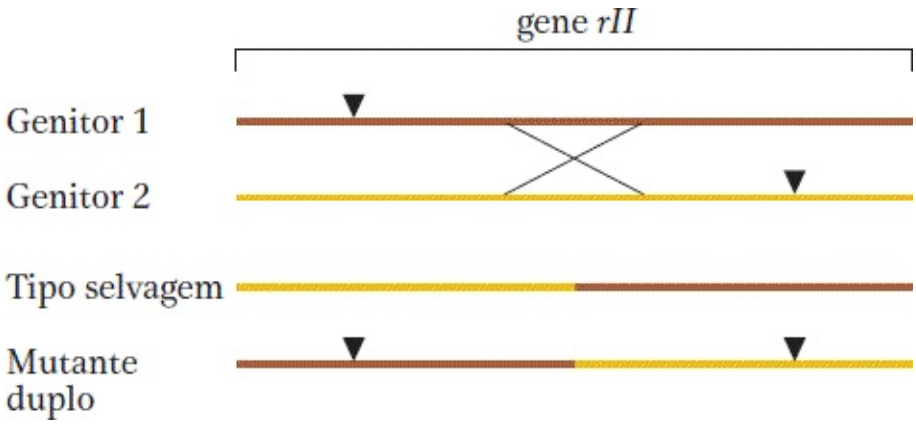


FIGURA 5.28 Estes fenótipos de placas foram produzidos pela progênie do cruzamento $h^{—}\ r^{+} \times h^{+}\ r^{—}$. Quatro fenótipos de placas podem ser diferenciados, representando dois tipos parentais e dois recombinantes. (De G. S. Stent, *Molecular Biology of Bacterial Viruses*. Direitos autorais 1963 por W. H. Freeman and Company.)

Tendo em vista que números astronomicamente grandes de fagos podem ser utilizados em análises de recombinação de fagos, podem ser detectados eventos de *crossover* muito raros. Na década de 1950, Seymour Benzer utilizou os referidos eventos de *crossover* raros para mapear os sítios mutantes no gene *rII* do fago T4, um gene que controla a lise. Em relação a diferentes alelos mutantes *rII* que surgem espontaneamente, o sítio mutante normalmente está em diferentes posições no gene. Portanto, quando dois mutantes *rII* diferentes são cruzados,

podem ocorrer alguns *crossovers* raros entre os sítios mutantes, produzindo recombinantes do tipo selvagem, conforme demonstrado aqui:



Na medida em que a distância entre dois sítios mutantes aumenta, um referido evento de *crossover* é mais provável. Portanto, a frequência de recombinantes rII^+ é uma medida dessa distância no gene. (O produto recíproco é um mutante duplo e indistinguível dos parentais.)

Benzer utilizou uma abordagem inteligente para detectar os muito raros recombinantes rII^+. Ele fez uso do fato de que os mutantes *rII* não infectarão uma linhagem de *E. coli* denominada K. Portanto, ele realizou o cruzamento *rII* × *rII* em outra linhagem e em seguida plaqueou o lisado de fagos em uma camada da linhagem K. Apenas os recombinantes rII^+ formarão placas nessa camada. Esse modo de observar um evento genético raro (nesse caso, um recombinante) é um **sistema seletivo**: *apenas* o evento raro desejado consegue produzir um determinado desfecho visível. Contrariamente, uma **triagem** é um sistema no qual grandes números de indivíduos são analisados visualmente para buscar a "agulha em um palheiro" rara.

Essa mesma abordagem pode ser utilizada para mapear sítios mutantes em genes de qualquer organismo a partir do qual grandes números de células possam ser obtidos e em relação ao qual os fenótipos do tipo selvagem e mutante possam ser distinguidos. Entretanto, esse tipo de mapeamento intragênico foi amplamente substituído pelo advento de métodos químicos não dispendiosos para sequenciamento do DNA, que identificam as posições dos sítios mutantes diretamente.

> CONCEITO-CHAVE A recombinação entre cromossomos de fagos pode ser estudada por meio da junção dos cromossomos parentais em uma célula hospedeira pela infecção mista. Os fagos da progênie podem ser examinados em relação aos genótipos parentais e recombinantes.

5.5 Transdução

Alguns fagos são capazes de captar genes bacterianos e carreá-los de uma célula bacteriana até outra, um processo conhecido como **transdução**. Portanto, a transdução junta-se à bateria de modos de transferência do material genômico entre bactérias — juntamente com a transferência do cromossomo Hfr, transferência do plasmídio F' e transformação.

Descoberta da transdução

Em 1951, Joshua Lederberg e Norton Zinder estavam testando a recombinação na bactéria *Salmonella typhimurium* por meio da utilização de técnicas que haviam apresentado sucesso com *E. coli*. Os pesquisadores utilizaram duas linhagens diferentes: uma era *phe*$^-$ *trp*$^-$ *tyr*$^-$ e a outra era *met*$^-$ *his*$^-$. Não nos preocuparemos com a natureza desses alelos, exceto para observar que todos são auxotróficos. Quando cada uma das linhagens foi plaqueada em um meio mínimo, não foram observadas células do tipo selvagem. Entretanto, após as duas linhagens serem misturadas, prototróficos do tipo selvagem apareceram em uma frequência de aproximadamente 1 em 10^5. Até então, a situação aparenta ser semelhante àquela em relação à recombinação em *E. coli*.

Entretanto, nesse caso, os pesquisadores também recuperaram recombinantes de um experimento com um tubo em formato de U, no qual a conjugação foi evitada por meio de um filtro que separava os dois braços (relembre a Figura 5.6). Eles formularam a hipótese de que algum agente carreava os genes de uma bactéria para outra. Ao variar o tamanho dos poros no filtro, eles observaram que o agente responsável pela transferência gênica era do mesmo tamanho de um fago de *Salmonella* conhecido, denominado fago P22. Além disso, o agente filtrável e

o P22 eram idênticos na sensibilidade ao antissoro e na imunidade às enzimas hidrolíticas. Portanto, Lederberg e Zinder haviam descoberto um novo tipo de transferência gênica, mediada por um vírus. Eles foram os primeiros a denominar esse processo *transdução*. Como uma raridade no ciclo lítico, as partículas de vírus por vezes captam genes bacterianos e os transferem quando infectam outro hospedeiro. A transdução foi demonstrada subsequentemente em muitas bactérias.

Para compreender o processo de transdução, precisamos distinguir dois tipos de ciclo de fagos. Os **fagos virulentos** são aqueles que lisam e matam o hospedeiro imediatamente. Os **fagos temperados** conseguem permanecer dentro da célula hospedeira durante um período sem matá-la. O seu DNA se integra ao cromossomo do hospedeiro para ser replicado com ele ou é replicado separadamente no citoplasma, assim como um plasmídio. Um fago integrado ao genoma bacteriano é denominado **profago**. Uma bactéria que abriga um fago quiescente é descrita como **lisogênica** e ela própria é denominada **lisógena**. Ocasionalmente, o fago quiescente em uma bactéria lisogênica se torna ativo, se replica e causa a lise espontânea da sua célula hospedeira. Um fago temperado residente confere resistência à infecção por outros fagos do mesmo tipo.

Existem dois tipos de transdução: generalizada e especializada. Os fagos de transdução *generalizada* conseguem carrear qualquer parte do cromossomo bacteriano, enquanto os fagos de transdução *especializada* carreiam apenas determinadas partes específicas.

> CONCEITO-CHAVE Os fagos virulentos não conseguem se tornar profagos; eles se replicam e lisam uma célula imediatamente. Os fagos temperados conseguem existir dentro da célula bacteriana como profagos, possibilitando que seus hospedeiros sobrevivam como bactérias lisogênicas; eles também são capazes de, ocasionalmente, lisar a bactéria.

Transdução generalizada

Por meio de quais mecanismos um fago consegue realizar a **transdução**

generalizada? Em 1965, H. Ikeda e J. Tomizawa lançaram luz sobre essa questão em alguns experimentos sobre o fago P1 de *E. coli*. Eles descobriram que, quando uma célula doadora é lisada por P1, o cromossomo bacteriano é rompido em pequenos fragmentos. Ocasionalmente, as partículas de fagos recentemente formadas incorporam, de modo errôneo, um fragmento do DNA bacteriano dentro de uma cabeça de fago, não no DNA do fago. Esse evento é a origem do fago transdutor.

Um fago que carreia o DNA bacteriano consegue infectar outra célula. Em seguida, aquele DNA bacteriano pode ser incorporado dentro do cromossomo da célula receptora por meio de recombinação (Figura 5.29). Tendo em vista que os genes em quaisquer das partes cortadas do genoma hospedeiro podem ser transduzidos, esse tipo de transdução é necessariamente do tipo generalizado.

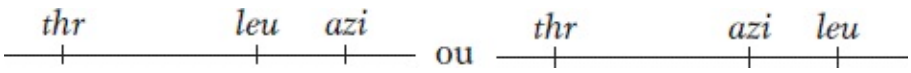
Ambos os fagos, P1 e P22, pertencem a um grupo de fagos que demonstra transdução generalizada. O DNA de P22 é inserido no cromossomo hospedeiro, enquanto o DNA de P1 permanece livre, como um grande plasmídio. Entretanto, ambos transduzem por meio do enchimento defeituoso da cabeça.

A transdução generalizada pode ser utilizada para a obtenção de informações sobre ligação bacteriana quando os genes estão suficientemente próximos para que o fago possa captá-los e transduzi-los em um único fragmento de DNA. Por exemplo, suponha que desejamos encontrar a distância de ligação entre *met* e *arg* em *E. coli*. Cultivamos o fago P1 em uma linhagem doadora met^+ arg^+ e em seguida possibilitamos que os fagos P1 da lise dessa linhagem infectem uma linhagem met^- arg^-. Primeiramente, é selecionado um alelo doador, digamos, met^+. Em seguida, é medida a porcentagem de colônias met^+ que também são arg^+. As linhagens transduzidas para ambos, met^+ e arg^+, são denominadas **cotransdutantes**. Quanto *maior* a frequência de cotransdução, *mais próximos* dois marcadores genéticos precisam estar (o oposto da maior parte das medidas de mapeamento). Os valores de ligação normalmente são expressos como frequências de cotransdução (Figura 5.30).

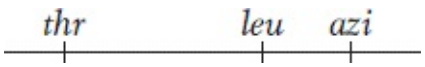
Ao utilizar uma extensão dessa abordagem, podemos estimar o *tamanho* do fragmento do cromossomo hospedeiro que um fago consegue captar, assim como no tipo de experimento a seguir, que utiliza o fago P1:

doador *leu*$^{+}$ *thr*$^{+}$ *azi*^{r} → receptor *leu*$^{-}$ *thr*$^{-}$ *azi*^{s}

Nesse experimento, o fago P1 cultivado na linhagem doadora *leu*$^{+}$ *thr*$^{+}$ *azi*^{r} infecta a linhagem receptora *leu*$^{-}$ *thr*$^{-}$ *azi*^{s}. A estratégia é selecionar um ou mais alelos doadores na receptora e em seguida testar esses transdutantes em relação à presença dos alelos não selecionados. Os resultados estão resumidos na Tabela 5.3. O experimento 1 na Tabela 5.3 nos informa que *leu* está relativamente próximo de *azi* e distante de *thr*, restando-nos duas possibilidades:



O experimento 2 nos informa que *leu* está mais próximo de *thr* do que *azi* está e, assim, o mapa tem de ser:



Ao selecionar *thr*$^{+}$ e *leu*$^{+}$ juntos nos fagos transdutores no experimento 3, observamos que o fragmento transduzido de material genético nunca inclui o *locus azi*, tendo em vista que a cabeça do fago não consegue carrear um fragmento de DNA tão grande quanto aquele. P1 somente consegue cotransduzir genes com distância inferior a aproximadamente 1,5 minuto no mapa cromossômico de *E. coli*.

Transdução generalizada por meio da incorporação aleatória de DNA bacteriano às cabeças de fagos

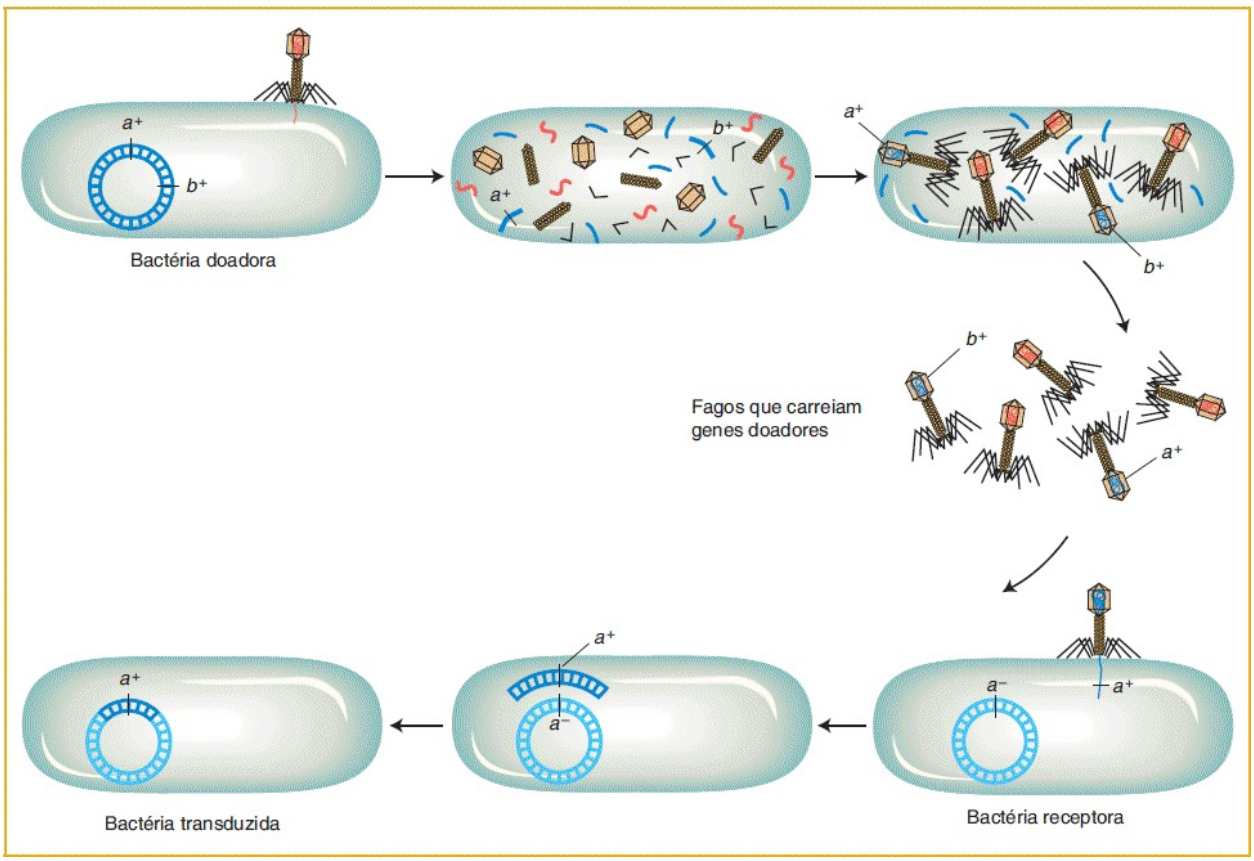


FIGURA 5.29 Um fago recentemente formado capta DNA do cromossomo de sua célula hospedeira (*parte superior*) e em seguida injeta-o em uma célula nova (*parte inferior à direita*). O DNA injetado insere-se no cromossomo do novo hospedeiro por meio de recombinação (*parte inferior à esquerda*). Na realidade, apenas uma minoria muito pequena da progênie de fagos (1 em 10.000) carreia genes doadores.

Ligação próxima é inferida a partir de altas frequências de cotransdução

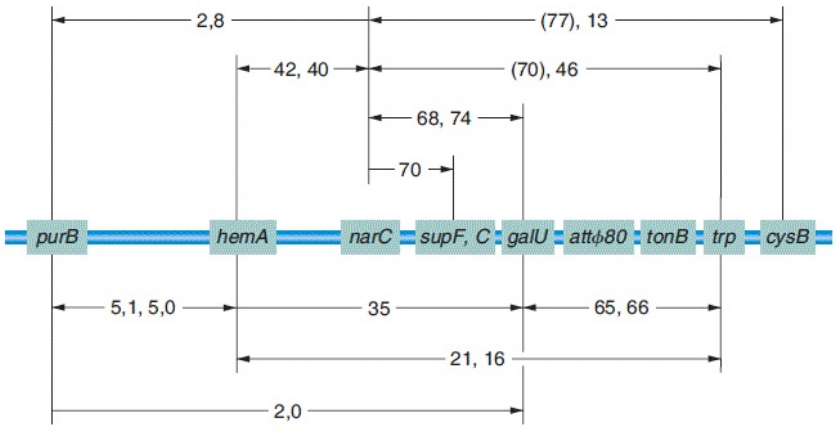


FIGURA 5.30 O diagrama demonstra um mapa genético da região de *purB* a *cysB* de *E. coli* determinada pela cotransdução de P1. Os números fornecidos são as médias em porcentagem em relação às frequências de transdução obtidas em diversos experimentos. Os valores entre parênteses são considerados não confiáveis. (Dados de J. R. Guest, *Mol. Gen. Genet.* 105, 1969, p. 285.)

Transdução especializada

Um transdutor generalizado, tal como o fago P22, capta fragmentos de DNA quebrado do hospedeiro aleatoriamente. Como os outros fagos, que atuam como transdutores especializados, conseguem carrear apenas determinados genes do hospedeiro até as células receptoras? A resposta curta é que um transdutor especializado insere-se no cromossomo bacteriano em apenas uma posição. Quando ele sai, ocorre uma alça defeituosa (semelhante ao tipo que produz plasmídios F'). Portanto, ele consegue captar e transduzir apenas genes que estão próximos.

Tabela 5.3 Marcadores de acompanhamento em transduções de P1 específicas.

Experimento	Marcador selecionado	Marcadores não selecionados
1	*leu*$^{+}$	50% são *azi*^{r}; 2% são *thr*$^{+}$
2	*thr*$^{+}$	3% são *leu*$^{+}$; 0% são *ari*^{r}
3	*leu*$^{+}$ e *thr*$^{+}$	0% são *azi*^{r}

O protótipo da **transdução especializada** foi fornecido por estudos realizados por Joshua e Esther Lederberg em um fago temperado de *E. coli* denominado *lambda* (λ). O fago λ se tornou o fago mais intensivamente estudado e mais bem caracterizado.

Comportamento do profago. O fago λ apresenta efeitos incomuns quando células lisogênicas em relação a ele são utilizadas em cruzamentos. No cruzamento de uma Hfr não infectada com uma receptora lisogênica F^- [Hfr × F^- (λ)], as exconjugantes F^- lisogênicas com genes Hfr são prontamente recuperadas. Entretanto, no cruzamento recíproco Hfr (λ) × F^-, os genes *iniciais* do cromossomo Hfr são recuperados entre as exconjugantes, mas os recombinantes em relação aos genes *tardios* não são recuperados. Além disso, exconjugantes lisogênicas quase nunca são recuperadas a partir desse cruzamento recíproco. Qual é a explicação? As observações fazem sentido se o profago λ estiver se comportando como um *locus* gênico bacteriano se comporta (ou seja, como parte do cromossomo bacteriano). Portanto, no cruzamento Hfr (λ) × F^-, o profago entraria na célula F^- em um momento específico, correspondente à sua posição no cromossomo. Os genes iniciais são recuperados porque eles entram antes do profago. Os genes tardios não são recuperados porque a lise destrói a célula receptora. Em experimentos de cruzamento interrompido, o profago λ de fato sempre entra na célula F^- em um momento específico, ligado de modo próximo ao *locus gal*.

Em um cruzamento Hfr (λ) × F^-, a entrada do profago λ na célula aciona imediatamente o profago para um ciclo lítico; esse processo é denominado **indução zigótica** (Figura 5.31). Entretanto, no cruzamento de *duas* células lisogênicas Hfr (λ) × F^- (λ), não existe indução zigótica. A presença de qualquer profago evita que outro vírus infectante cause lise. Isso porque o profago produz um fator citoplasmático que reprime a multiplicação do vírus. (O repressor citoplasmático fago-direcionado explica muito bem a imunidade das bactérias lisogênicas, tendo em vista que um fago encontraria imediatamente um repressor e seria inativado.)

Inserção do λ. Os experimentos de cruzamento interrompido até aqui descritos demonstraram que o profago λ faz parte do cromossomo da bactéria lisogênica. Como o profago λ é inserido no genoma bacteriano? Em 1962, Allan Campbell propôs que ele é inserido por meio de um *crossover* único entre um cromossomo de fago λ circular e o cromossomo de *E. coli* circular, conforme demonstrado na

Figura 5.32. O ponto do *crossover* estaria entre um sítio específico em λ, o **sítio de ligação do** λ, e um sítio de ligação no cromossomo bacteriano, localizado entre os genes *gal* e *bio*, tendo em vista que λ integra-se naquela posição no cromossomo de *E. coli*.

Uma atração da proposta de Campbell é que a partir dela se seguem previsões que os geneticistas conseguem testar. Por exemplo, a integração do profago ao cromossomo de *E. coli* deve aumentar a distância genética entre os genes bacterianos flanqueadores, conforme pode ser observado na Figura 5.32 em relação a *gal* e *bio*. De fato, estudos demonstram que a lisogenia *aumenta* o tempo de entrada ou as distâncias de recombinação entre os genes bacterianos. Essa localização única de λ contribui para sua transdução especializada.

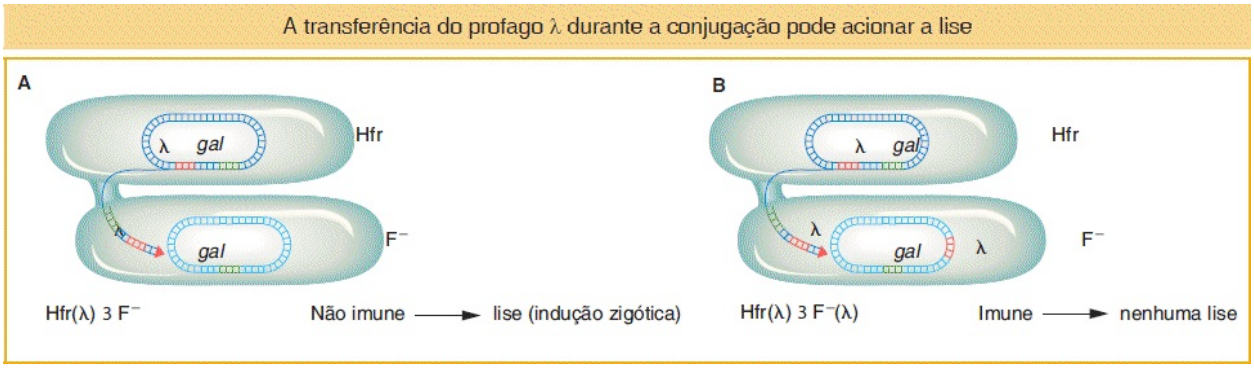


FIGURA 5.31 Um profago λ pode ser transferido para uma receptora durante a conjugação, mas o profago aciona a lise, um processo denominado indução zigótica, apenas se a receptora já não apresentar profago — ou seja, no caso demonstrado na parte **A**, mas não na parte **B**.

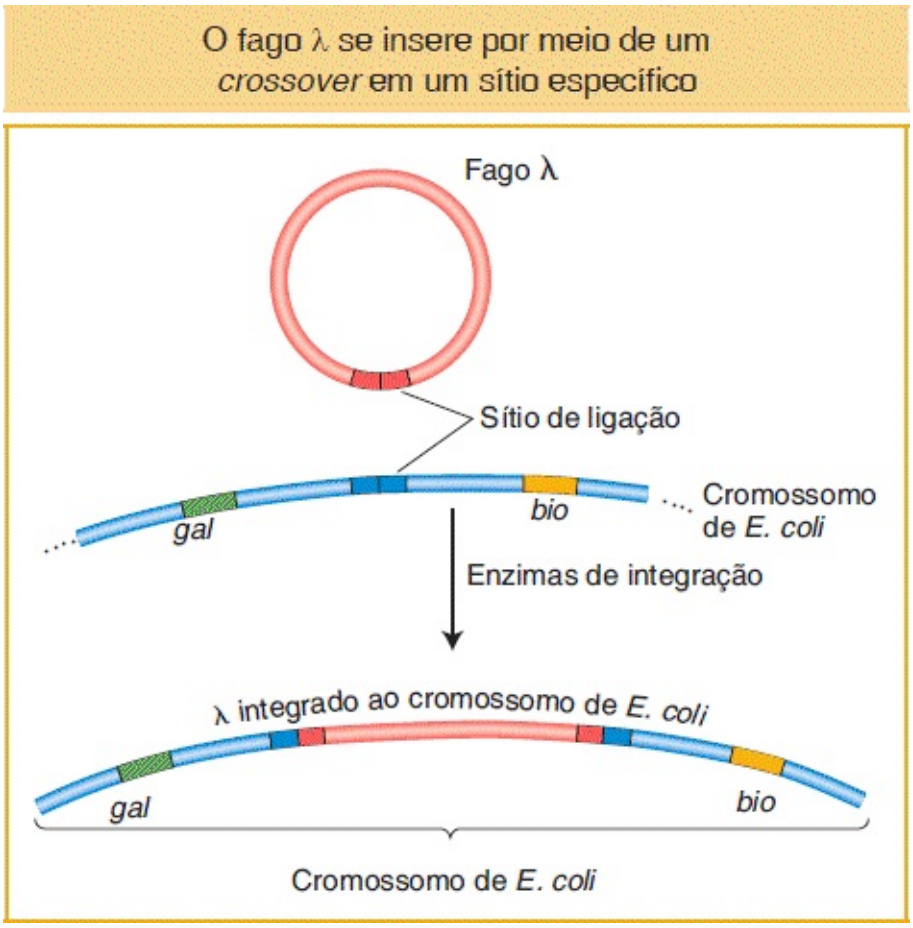


FIGURA 5.32 A recombinação recíproca ocorre entre um sítio de ligação específico no DNA circular e uma região específica denominada sítio de ligação no cromossomo de *E. coli* entre os genes *gal* e *bio*.

Mecanismo de transdução especializada

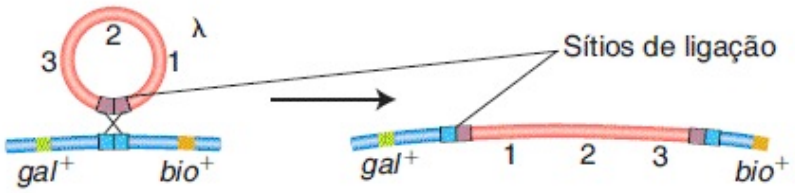
Como é um profago, λ sempre se insere entre a região *gal* e a região *bio* do cromossomo hospedeiro (Figura 5.33) e, em experimentos de transdução, conforme esperado, λ consegue transduzir apenas os *genes gal* e *bio*.

Como λ transporta os genes vizinhos? Novamente, a explicação está em uma reversão imperfeita do mecanismo de inserção de Campbell, como aquele em relação à formação de F'. O evento de recombinação entre regiões específicas de λ e o cromossomo bacteriano é catalisado por um sistema enzimático especializado codificado pelo fago, que utiliza o sítio de ligação de λ como um substrato. O sistema enzimático determina que λ integre-se apenas em um ponto específico entre *gal* e *bio* no cromossomo (ver Figura 5.33 A). Além disso,

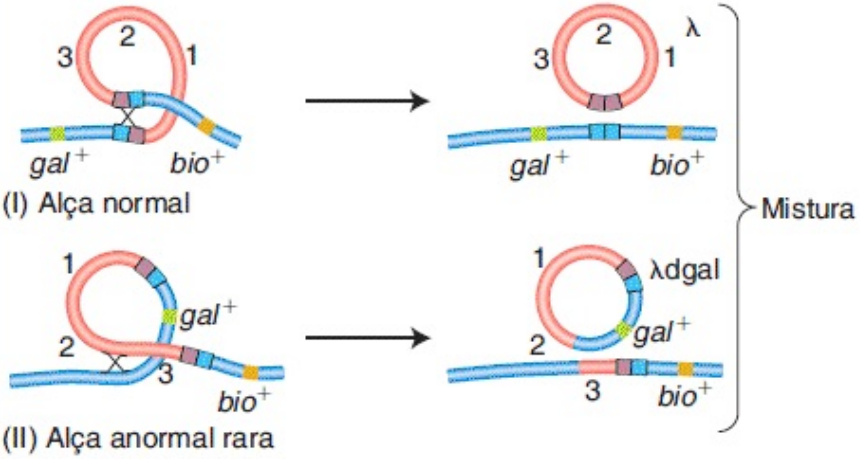
durante a lise, o profago λ normalmente é excisado precisamente no ponto correto para produzir um cromossomo λ circular normal, conforme observado na Figura 5.33 B (I). Muito raramente, a excisão é anormal em virtude da alça defeituosa. Nesse caso, a alça de DNA do fago pode captar um gene próximo e deixar para trás alguns genes do fago, conforme observado na Figura 5.33 B (II). O genoma do fago resultante é defeituoso em virtude dos genes deixados para trás, mas também obteve um gene bacteriano, *gal* ou *bio*. O DNA anormal que carreia os genes próximos pode ser condensado nas cabeças dos fagos para produzir partículas de fagos que conseguem infectar outras bactérias. Esses fagos são denominados λdgal (*gal* λ-defeituoso) ou λdbio. Na presença de uma segunda partícula de fago normal em uma infecção dupla, λdgal consegue integrar-se ao cromossomo no sítio de ligação de λ (Figura 5.33 C). Desse modo, os genes *gal* nesse caso são transduzidos para o segundo hospedeiro.

Alça defeituosa produz fago λ que contém DNA bacteriano

A. Produção de lisógeno



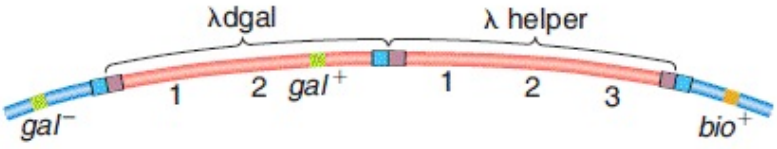
B. Produção de lisado inicial



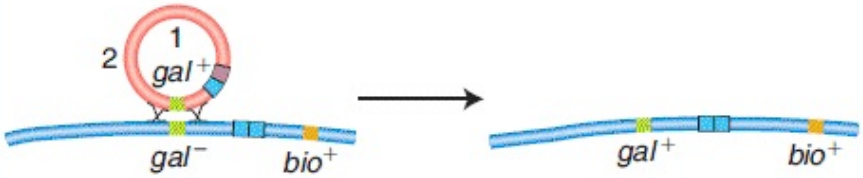
(I) Alça normal

(II) Alça anormal rara

C. Transdução por lisado inicial



(I) Transdutantes lisogênicos



(II) Transdutantes produzidos por recombinação

FIGURA 5.33 O diagrama demonstra como a transdução especializada é operada no fago λ. **A.** Um *crossover* no sítio de ligação especializado produz uma bactéria lisogênica. **B.** A bactéria lisogênica pode produzir um λ normal (I), ou, raramente, λdgal (II), uma partícula transdutora que contém o gene *gal*. **C.** Transdutantes gal^+ podem ser produzidos por meio (I) da coincorporação de λdgal e λ (que atua como um auxiliar) ou (II) de *crossovers* flanqueadores do gene *gal*, um evento raro. Os quadrados duplos azuis são sítios de ligação bacteriana, os quadrados duplos roxos são o sítio de ligação de λ e os pares dos quadrados azuis e roxos são os sítios híbridos de integração, derivados parcialmente de *E. coli* e parcialmente de λ.

CONCEITO-CHAVE A transdução ocorre quando fagos recentemente formados adquirem genes do hospedeiro e os transferem para outras células bacterianas. A *transdução generalizada* consegue transferir qualquer gene do hospedeiro. Ela ocorre quando o processo de condensação do fago acidentalmente incorpora DNA bacteriano em vez de DNA do fago. A *transdução especializada* ocorre em virtude da alça defeituosa do profago do cromossomo bacteriano e, assim, o novo fago contém os genes do fago e bacterianos. O fago transdutor só consegue transferir genes específicos do hospedeiro.

5.6 Comparação de mapas físicos e mapas de ligação

Alguns mapas cromossômicos muito detalhados de bactérias foram obtidos por meio da combinação das técnicas de mapeamento de cruzamento interrompido, mapeamento de recombinação, transformação e transdução. Atualmente, novos marcadores genéticos são tipicamente mapeados primeiramente em um segmento de aproximadamente 10 a 15 minutos de mapa por meio da utilização do cruzamento interrompido. Em seguida, marcadores adicionais ligados de modo próximo podem ser mapeados em uma análise mais meticulosa com a utilização de cotransdução de P1 ou recombinação.

Em 1963, o mapa de *E. coli* (Figura 5.34) já detalhava as posições de aproximadamente 100 genes. Após 27 anos de refinamento adicional, o mapa de 1990 ilustrava as posições de mais de 1.400 genes. A Figura 5.35 demonstra uma seção de 5 minutos do mapa de 1990 (que está ajustado para uma escala de 100

minutos). A complexidade desses mapas ilustra o poder e a sofisticação da análise genética. Quão bem esses mapas correspondem à realidade física? Em 1997, a sequência do DNA de todo o genoma de *E. coli* de 4.632.221 pares de bases foi concluída, o que nos possibilita comparar a posição exata dos genes no mapa genético com a posição da sequência codificadora correspondente na sequência do DNA linear (o mapa físico). O mapa integral está representado na Figura 5.36. A Figura 5.37 realiza uma comparação em relação a um segmento de ambos os mapas. Claramente, o mapa genético é uma correspondência próxima do mapa físico.

Um mapa do genoma de *E. coli* obtido geneticamente

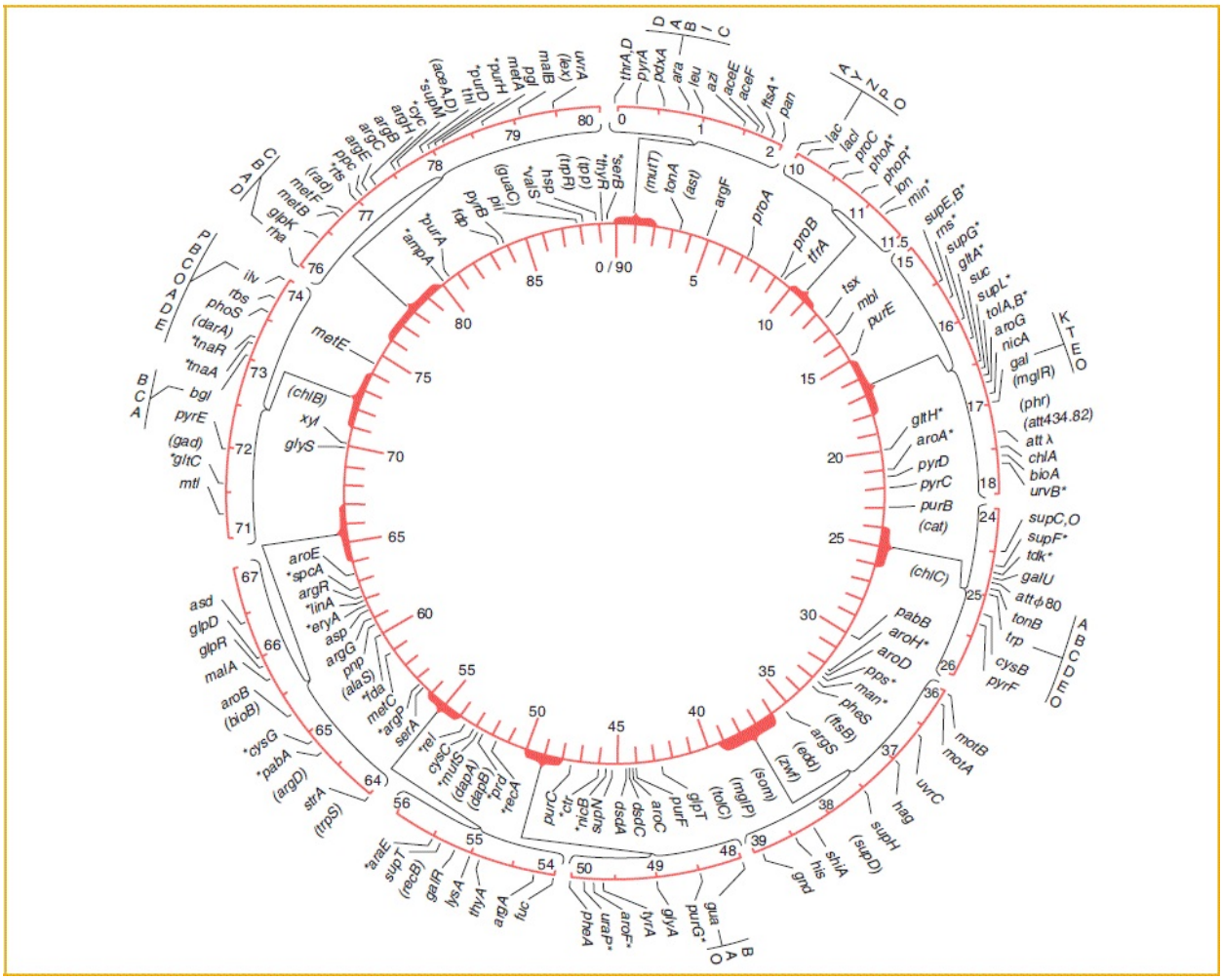


FIGURA 5.34 O mapa genético de 1963 dos genes de *E. coli*, com fenótipos mutantes. As unidades são minutos, com base em experimentos de cruzamento interrompido e recombinação. Os asteriscos fazem

referência às posições do mapa que não são tão precisas quanto as outras posições. (Dados de G. S. Stent, *Molecular Biology of Bacterial Viruses*.)

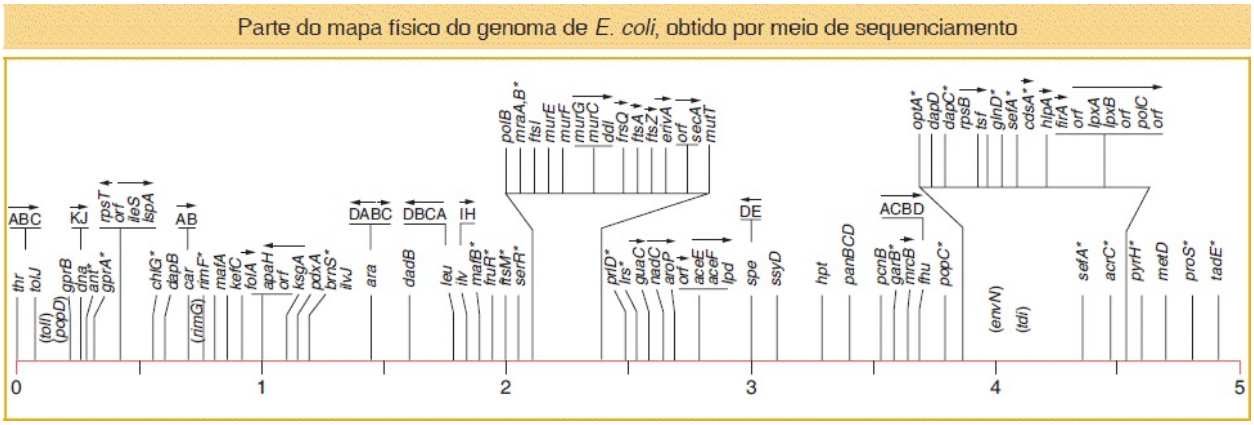


FIGURA 5.35 Um desenho em escala linear de uma seção de 5 min sequenciada do mapa de ligação de *E. coli* de 1990 de 100 min. Os parênteses e os asteriscos indicam marcadores em relação aos quais a localização exata era desconhecida na ocasião da publicação. As *setas* acima dos genes e dos grupos de genes indicam o sentido da transcrição. (Dados de B. J. Bachmann, "Linkage Map of Escherichia coli K-12, Edition 8", *Microbiol. Rev.* 54, 1990, 130-197.)

O Capítulo 4 considerou alguns modos pelos quais o mapa físico (normalmente a sequência completa do genoma) pode ser útil no mapeamento de novas mutações. Em bactérias, a técnica da **mutagênese insercional** é outro modo de detectar rapidamente a posição de uma mutação em um mapa físico conhecido. A técnica causa mutações por meio da inserção aleatória de fragmentos de DNA "estranhos". Os insertos inativam qualquer gene no qual penetrem por meio da interrupção da unidade transcricional. Os transpósons são inserções particularmente úteis para essa finalidade em diversos organismos-modelo, incluindo bactérias. Para mapear uma nova mutação, o procedimento é como segue. O DNA de um transpóson que carreia um alelo de resistência ou outro marcador selecionável é introduzido por meio de transformação nas receptoras bacterianas que não apresentam transpósons ativos. Os transpósons inserem-se mais ou menos aleatoriamente e qualquer um que entre no meio de um gene causa uma mutação. Um subconjunto de todos os mutantes obtidos apresentará fenótipos relevantes para o processo bacteriano em estudo e esses fenótipos se tornam o foco da análise.

A beleza da inserção dos transpósons é que, tendo em vista que sua sequência é conhecida, o gene mutante pode ser localizado e sequenciado. São criados *primers* para a replicação do DNA que correspondem à sequência conhecida do transpóson (ver Capítulo 10). Esses *primers* são utilizados para iniciar uma análise de sequenciamento que ocorre de modo centrífugo a partir do transpóson *em direção ao* gene adjacente. Em seguida, a curta sequência obtida pode ser colocada em um computador e comparada à sequência completa do genoma. A partir dessa análise, a posição do gene e sua sequência completa são obtidas. A função de um homólogo desse gene pode já ter sido deduzida em outros organismos. Portanto, você pode verificar que essa abordagem (assim como aquela introduzida no Capítulo 4) é outro modo de unir o fenótipo mutante com a posição do mapa e a possível função. A Figura 5.38 resume a abordagem.

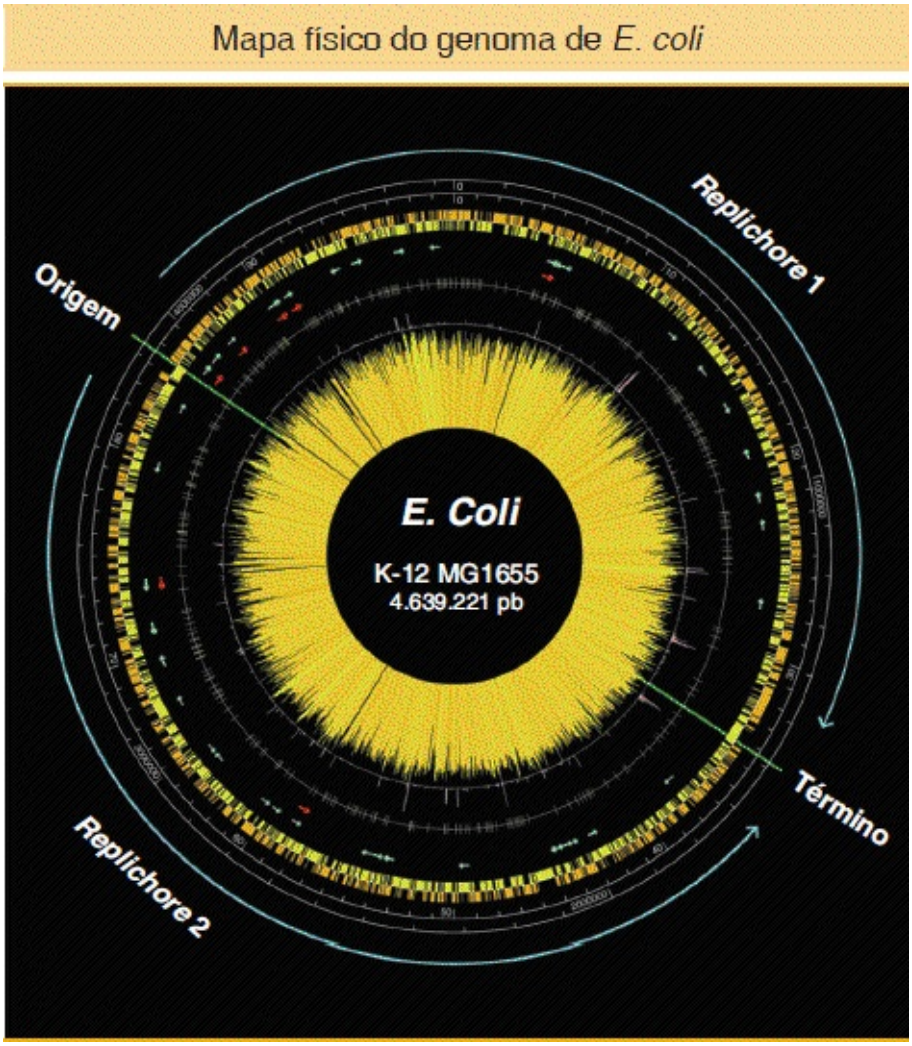


FIGURA 5.36 Este mapa foi obtido a partir do sequenciamento do DNA e da plotagem das posições dos genes. Legenda dos componentes, do exterior para o interior:

- A origem e o término da replicação do DNA estão marcados
- As duas escalas estão em pares de base de DNA e em minutos
- Os histogramas em *laranja* e *amarelo* demonstram a distribuição dos genes nos dois filamentos de DNA diferentes
- As *setas* representam os genes para rRNA (*vermelhas*) e tRNA (*verdes*)
- A "explosão estelar" central é um histograma de cada gene, com linhas de comprimento que refletem o nível de transcrição previsto.

(F. R. Blattner *et al.*, "The Complete Genome Sequence of Escherichia coli K-12", *Science* 277, 1997, 1453-1462. DOI: 10.1126/science.277.5331.1453. Reproduzida, com autorização, da AAAS. Cortesia do Dr. Guy Plunkett III.)

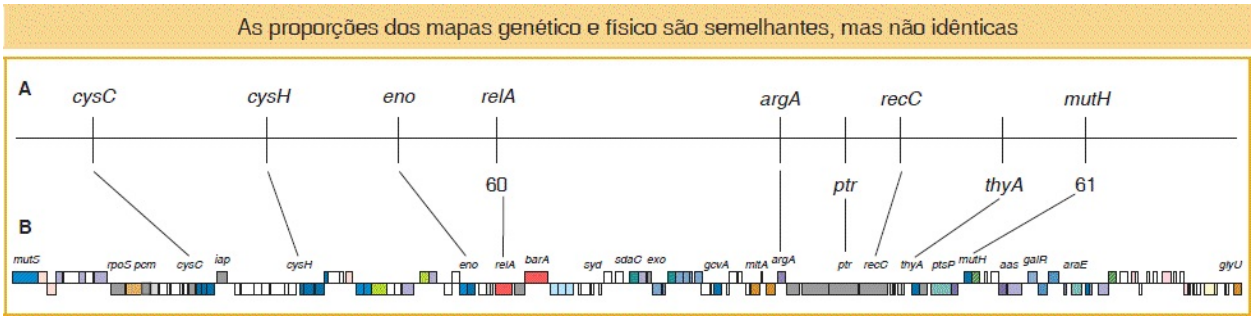


FIGURA 5.37 Um alinhamento dos mapas genético e físico. **A.** Marcadores no mapa genético de 1990 na região próxima de 60 e 61 min. **B.** As posições exatas de cada gene, com base na sequência completa do genoma de *E. coli*. (Nem todo gene está denominado neste mapa, para simplicidade.) Os quadros alongados são genes e genes putativos. Cada cor representa um tipo de função diferente. Por exemplo, *vermelho* indica funções regulatórias e *azul-escuro* indica funções na replicação, na recombinação e no reparo do DNA. As linhas entre os mapas nas partes **A** e **B** conectam o mesmo gene em cada mapa. (Dados de F. R. Blattner *et al.*, *The Complete Science* 277, 1997, 1453-1462.)

É interessante mencionar que muitos dos experimentos históricos que revelam a circularidade dos genomas bacterianos e de plasmídios coincidiram com a publicação e a popularização de *O Senhor dos Anéis*, de J. R. R. Tolkien. Consequentemente, uma revisão da genética bacteriana naquela ocasião começou com a seguinte citação da trilogia:

Um Anel para a todos governar, um Anel para encontrá-los,

Um Anel para a todos trazer e na escuridão aprisioná-los.

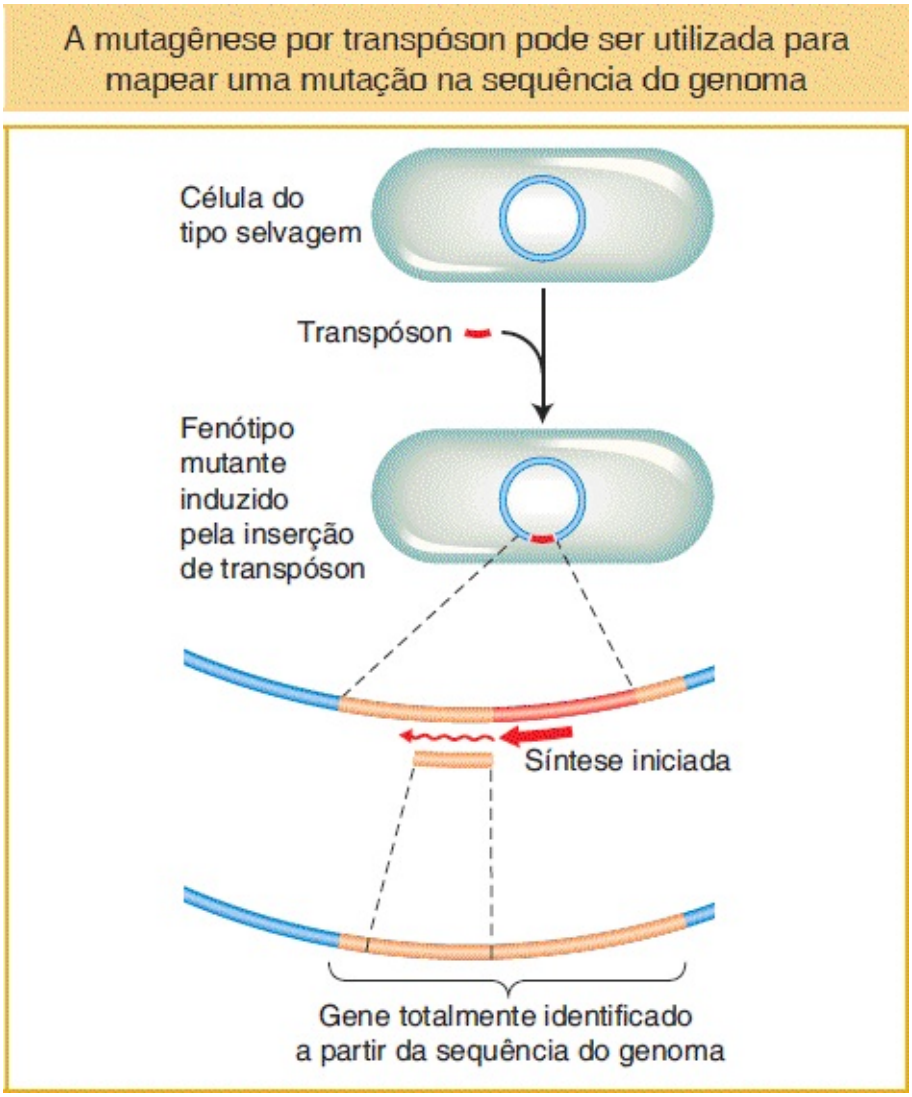


FIGURA 5.38 A inserção de um transpóson insere uma mutação em um gene de posição e função desconhecidas. O segmento próximo do transpóson é replicado, sequenciado e corresponde a um segmento na sequência completa do genoma.

RESUMO

Avanços na genética bacteriana e de fagos nos últimos 50 anos proporcionaram o fundamento para a biologia molecular e a clonagem (discutida nos capítulos posteriores). Inicialmente nesse período, observou-se que a transferência e a recombinação de genes ocorrem entre diferentes linhagens de bactérias.

Entretanto, nas bactérias, o material genético é transmitido em apenas um sentido — por exemplo, na *Escherichia coli*, de uma célula doadora (F^+ ou Hfr) para uma célula receptora (F^-). A capacidade do doador é determinada pela presença, na célula, de um fator de fertilidade (F), um tipo de plasmídio. Às vezes, o fator F presente no estado livre nas células F^+ consegue integrar-se ao cromossomo de *E. coli* e formar uma célula Hfr. Quando isso ocorre, um fragmento do cromossomo doador pode ser transferido para uma célula receptora e subsequentemente recombinar-se com o cromossomo receptor. Tendo em vista que o fator F pode ser inserido em diferentes locais no cromossomo hospedeiro, os investigadores iniciais conseguiram reunir os fragmentos transferidos para demonstrar que o cromossomo de *E. coli* é um círculo único, ou anel. A interrupção da transferência em tempos diferentes proporcionou aos geneticistas um método não convencional (cruzamento interrompido) para a construção de um mapa de ligação do único cromossomo de *E. coli* e de outras bactérias semelhantes, no qual a unidade de mapa é uma unidade de tempo (minutos). Em uma extensão dessa técnica, a frequência de recombinantes entre marcadores que sabidamente entraram na receptora pode determinar melhor uma distância de mapa.

Podem ser observados diversos tipos de plasmídios além de F. Os plasmídios R carreiam alelos de resistência a antibióticos, com frequência em um elemento móvel denominado transpóson. A rápida difusão dos plasmídios causa uma resistência ampla na população a fármacos de importância médica. Os derivados dos referidos plasmídios naturais se tornaram importantes vetores de clonagem, úteis para o isolamento dos genes e estudo em todos os organismos.

Os traços genéticos também podem ser transferidos de uma célula bacteriana para outra na forma de fragmentos de DNA transferidos para dentro da célula a partir do ambiente extracelular. Esse processo de transformação nas células bacterianas foi a primeira demonstração de que o DNA é o material genético. Para que ocorra a transformação, o DNA tem de ser transferido para uma célula receptora, e em seguida é necessário ocorrer recombinação entre o cromossomo receptor e o DNA incorporado.

As bactérias podem ser infectadas por vírus denominados bacteriófagos. Em um método de infecção, o cromossomo do fago penetra na célula bacteriana e

utiliza o maquinário metabólico bacteriano para produzir fagos que rompem a bactéria hospedeira. Em seguida os novos fagos infectam outras células. Se dois fagos de genótipos diferentes infectarem a mesma hospedeira, pode ocorrer recombinação entre os seus cromossomos.

Em outro modo de infecção, a lisogenia, o fago injetado permanece quiescente na célula bacteriana. Em muitos casos, esse fago quiescente (o profago) incorpora-se ao cromossomo da hospedeira e é replicado com ele. Seja espontaneamente ou sob a estimulação apropriada, o profago pode deixar seu estado quiescente e lisar a célula hospedeira bacteriana.

Um fago pode carrear genes bacterianos de um doador para um receptor. Na transdução generalizada, o DNA aleatório do hospedeiro é incorporado isoladamente dentro da cabeça do fago durante a lise. Na transdução especializada, a excisão defeituosa do profago de um *locus* cromossômico único resulta na inclusão de genes específicos do hospedeiro, bem como de DNA do fago na cabeça do fago.

Atualmente, um mapa físico na forma da sequência completa do genoma está disponível em relação a muitas espécies bacterianas. Com a utilização desse mapa físico do genoma, a posição no mapa de uma mutação de interesse pode ser localizada com precisão. Primeiramente, são produzidas mutações apropriadas por meio da inserção de transpósons (mutagênese insercional). Em seguida, a sequência de DNA adjacente ao transpóson inserido é obtida e correlacionada com uma sequência no mapa físico. Essa técnica fornece o *locus*, a sequência e, possivelmente, a função do gene de interesse.

TERMOS-CHAVE

auxotrófico
bacteriófago (fago)
clone celular
colônia
conjugação
cotransdutante

cruzamento interrompido
doadora
endogenoto
exconjugante
exogenoto
F^+ (doadora)
F^- (receptora)
fago (bacteriófago)
fago temperado
fago virulento
fator de fertilidade (F)
Hfr (alta frequência de recombinação)
indução zigótica
infecção dupla (mista)
infecção mista (dupla)
lisado
lise
lisógena (bactéria lisogênica)
marcador genético
marcador não selecionado
meio mínimo
merozigoto
mutagênese insercional
mutante resistente
origem (O)
placa
plaqueamento
plasmídio
plasmídio F'
plasmídio R
procarioto
profago

prototrófica
receptora
recombinação de fago
replicação por círculo rolante
sistema seletivo
sítio de ligação de λ
término
transdução
transdução especializada
transdução generalizada
transformação
transformação dupla
transmissão horizontal
transmissão vertical
triagem
vírus

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Suponha que uma célula seja incapaz de realizar a recombinação generalizada (*rec*$^{-}$). Como essa célula se comportaria como uma receptora na transdução generalizada e na transdução especializada? Primeiramente, compare cada tipo de transdução, e em seguida determine o efeito da mutação *rec*$^{-}$ sobre a herança dos genes por meio de cada processo.

Solução

A transdução generalizada envolve a incorporação de fragmentos cromossômicos às cabeças dos fagos, que em seguida infectam as linhagens receptoras. Os fragmentos do cromossomo são incorporados aleatoriamente às cabeças dos fagos e, assim, qualquer marcador no cromossomo hospedeiro bacteriano pode ser transduzido para outra linhagem por meio da transdução generalizada. Contrariamente, a transdução especializada envolve a integração do fago em um

ponto específico no cromossomo e a rara incorporação dos marcadores cromossômicos próximo ao sítio de integração no genoma do fago. Portanto, apenas aqueles marcadores que estão próximos do sítio de integração específico do fago no cromossomo hospedeiro podem ser transduzidos.

Os marcadores são herdados por vias diferentes na transdução generalizada e na transdução especializada. Um fago de transdução generalizada injeta um fragmento do cromossomo doador no receptor. Esse fragmento tem de ser incorporado ao cromossomo do receptor por recombinação, com a utilização do sistema de recombinação do receptor. Portanto, um receptor *rec*$^-$ não será capaz de incorporar os fragmentos de DNA e não consegue herdar marcadores por meio da transdução generalizada. Por outro lado, a principal via de herança de marcadores por transdução especializada é por meio da integração da partícula de transdução especializada no cromossomo hospedeiro no sítio específico de integração do fago. Essa integração, que por vezes requer um fago do tipo selvagem adicional (auxiliar), é mediada por um sistema enzimático específico do fago, que é independente das enzimas de recombinação normais. Portanto, um receptor *rec*$^-$ ainda pode herdar marcadores genéticos por meio da transdução especializada.

Problema resolvido 2. Em *E. coli*, quatro linhagens Hfr doam os marcadores genéticos a seguir, demonstrados na ordem doada:

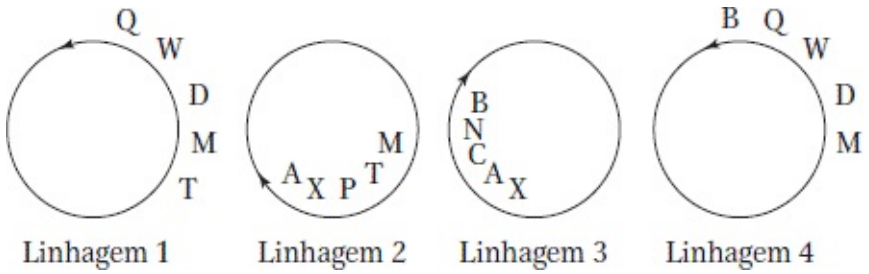
Linhagem 1:	Q	W	D	M	T
Linhagem 2:	A	X	P	T	M
Linhagem 3:	B	N	C	A	X
Linhagem 4:	B	Q	W	D	M

Todas essas linhagens Hfr são derivadas da mesma linhagem F^+. Qual é a ordem desses marcadores no cromossomo circular da F^+ original?

Solução

Uma abordagem de duas etapas funciona bem: (1) determine o princípio subjacente e (2) desenhe um diagrama. Aqui, o princípio claramente é que cada

linhagem Hfr doa marcadores genéticos a partir de um ponto fixo no cromossomo circular e que os marcadores mais iniciais são doados com mais alta frequência. Tendo em vista que nem todos os marcadores são doados por cada Hfr, apenas os marcadores iniciais tem de ser doados para cada Hfr. Cada linhagem nos possibilita desenhar os círculos a seguir:



A partir dessas informações, podemos consolidar cada círculo em um mapa de ligação circular na ordem Q, W, D, M, T, P, X, A, C, N, B, Q.

Problema resolvido 3. Em um cruzamento Hfr × F⁻, *leu*⁺ entra como o primeiro marcador, mas a ordem dos outros marcadores é desconhecida. Se Hfr for do tipo selvagem e F⁻ for auxotrófica em relação a cada marcador em questão, qual é a ordem dos marcadores em um cruzamento no qual recombinantes *leu*⁺ são selecionados, se 27% forem *ile*⁺, 13% forem *mal*⁺, 82% forem *thr*⁺, e 1% for *trp*⁺?

Solução

Relembre que a quebra espontânea cria um gradiente de transferência natural, o qual torna a cada vez menos provável que uma receptora receba marcadores cada vez mais tardios. Tendo em vista que selecionamos o marcador mais inicial nesse cruzamento, a frequência de recombinantes depende da ordem de entrada em relação a cada marcador. Portanto, podemos determinar imediatamente a ordem dos marcadores genéticos simplesmente ao observar a porcentagem de recombinantes em relação a qualquer marcador entre os recombinantes *leu*⁺. Tendo em vista que a herança de *thr*⁺ é a mais alta, *thr*⁺ tem de ser o primeiro marcador a entrar após *leu*. A ordem completa é *leu, thr, ile, mal, trp.*

Problema resolvido 4. É realizado um cruzamento entre uma Hfr que é *met*$^+$ *thi*$^+$ *pur*$^+$ e uma F$^-$ que é *met*$^-$ *thi*$^-$ *pur*$^-$. Estudos de cruzamento interrompido demonstram que *met*$^+$ entra no receptor por último e, assim, recombinantes *met*$^+$ são selecionados em um meio que contém suplementos que atendem apenas as exigências de *pur* e *thi*. Esses recombinantes são testados à procura dos alelos *thi*$^+$ e *pur*$^+$. Os números de indivíduos a seguir são observados em relação a cada genótipo:

met$^+$ *thi*$^+$ *pur*$^+$	280
met$^+$ *thi*$^+$ *pur*$^-$	0
met$^+$ *thi*$^-$ *pur*$^+$	6
met$^+$ *thi*$^-$ *pur*$^-$	52

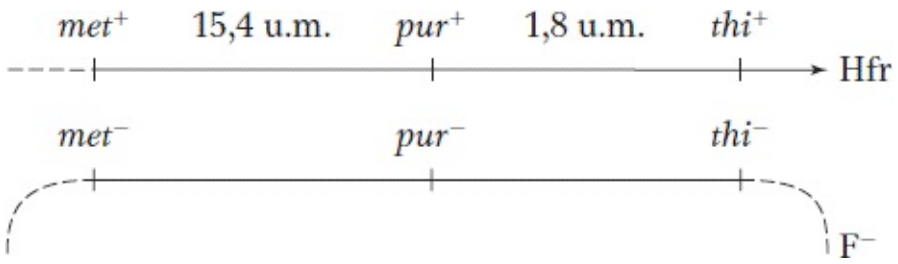
a. Por que a metionina (met) foi deixada fora do meio de seleção?
b. Qual é a ordem dos genes?
c. Quais são as distâncias de mapa em unidades de recombinação?

Solução

a. A metionina foi deixada fora do meio para possibilitar a seleção em relação aos recombinantes *met*$^+$, tendo em vista que *met*$^+$ é o último marcador a entrar na receptora. A seleção em relação a *met*$^+$ assegura que todos os *loci* que estamos considerando no cruzamento já terão entrado em cada recombinante que analisamos.
b. Aqui, um diagrama das possíveis ordens de genes é útil. Tendo em vista que sabemos que *met* entra na receptora por último, existem apenas duas ordens de genes possíveis se o primeiro marcador entrar à direita: *met, thi, pur* ou *met, pur, thi*. Como podemos distinguir entre essas duas ordens? Felizmente, uma das quatro possíveis classes de recombinantes requer dois *crossovers* adicionais. Cada ordem possível prevê uma classe diferente que surge por meio de quatro *crossovers*, em vez de dois. Por exemplo, se a ordem fosse *met, thi, pur*, os recombinantes *met*$^+$ *thi*$^-$ *pur*$^+$ seriam muito raros. Por outro lado, se a ordem fosse *met, pur, thi*, a classe de quatro *crossovers* seria *met*$^+$ *pur*$^-$ *thi*$^+$. A partir

das informações fornecidas na tabela, a classe *met*$^+$ *pur*$^-$ *thi*$^+$ é claramente a classe de quatro *crossovers* e, portanto, a ordem de genes *met*, *pur*, *thi* está correta.

c. Consulte o diagrama a seguir:



Para computar a distância entre *met* e *pur*, calculamos a porcentagem de *met*$^+$ *pur*$^-$ *thi*$^-$, que é 52/338 = 15,4 u.m. De modo semelhante, a distância entre *pur* e *thi* é 6/338 = 1,8 u.m.

Problema resolvido 5. Compare o mecanismo de transferência e herança dos genes *lac*$^+$ em cruzamentos com linhagens Hfr, F$^+$, e F' *lac*$^+$. Como uma célula F$^-$ que não pode sofrer recombinação homóloga normal (*rec*$^-$) se comporta em cruzamentos com cada uma dessas três linhagens? Como a célula seria capaz de herdar o gene *lac*$^+$?

Solução

Cada uma dessas três linhagens doa genes por meio de conjugação. Nas linhagens Hfr e F$^+$, os genes *lac*$^+$ no cromossomo hospedeiro são doados. Na linhagem Hfr, o fator F é integrado ao cromossomo em todas as células e, assim, os marcadores cromossômicos podem ser doados com eficiência, em particular se um marcador estiver próximo do sítio de integração de F e for doado inicialmente. A população de células F$^+$ contém uma pequena porcentagem de células Hfr, nas quais F é integrado ao cromossomo. Essas células são responsáveis pela transferência de genes ilustrada por culturas de células F$^+$. Na transferência de genes mediada por Hfr$^-$ e F$^+$, a herança requer a incorporação de um fragmento transferido por recombinação (relembre que são necessários dois *crossovers*) no cromossomo da F$^-$. Portanto, uma linhagem F$^-$ que não consegue sofrer recombinação não pode

herdar marcadores cromossômicos da doadora, embora eles sejam transferidos por linhagens Hfr ou por células Hfr em linhagens F^+. O fragmento não pode ser incorporado ao cromossomo por recombinação. Tendo em vista que esses fragmentos não possuem a capacidade de se replicar na célula F^-, eles são diluídos rapidamente durante a divisão celular.

Contrariamente às células Hfr, as células F' transferem os genes carreados pelo fator F', um processo que não requer a transferência do cromossomo. Nesse caso, os genes *lac*$^+$ estão ligados ao fator F' e são transferidos com ele a uma alta eficiência. Na célula F^-, nenhuma recombinação é necessária, tendo em vista que a linhagem F' *lac*$^+$ consegue replicar-se e ser mantida na população de células F^- em divisão. Portanto, os genes *lac*$^+$ são herdados até mesmo em uma linhagem *rec*$^-$.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 5.2, em qual dos quatro processos demonstrados um genoma bacteriano completo pode ser transferido de uma célula para outra?

2. Na Figura 5.3, se a concentração de células bacterianas na suspensão original for 200/mℓ e se 0,2 mℓ for plaqueado em cada uma de 100 placas de Petri, qual é a quantidade média esperada de colônias por placa?

3. Na Figura 5.5,

a. Por que as células A^- e B^-, por si próprias, não formam colônias no meio da cultura na placa?

b. Qual evento genético as colônias roxas na placa do meio representam?

4. Na Figura 5.10 C, o que os pontos amarelos representam?

5. Na Figura 5.11, quais alelos doadores se tornam parte do genoma recombinante produzido?

6. Na Figura 5.12,

a. Qual gene Hfr entra na receptora por último? (Qual diagrama o demonstra realmente entrando?)

b. Qual é a porcentagem máxima de casos de transferência desse gene?

c. Quais genes entraram em 25 minutos? Eles podem todos ter se tornado parte de um genoma exconjugante estável?

7. Na Figura 5.14, qual é o último gene a ser transferido para a F^- a partir de cada uma das cinco linhagens Hfr?

8. Na Figura 5.15, como cada um dos genótipos a seguir é produzido?

a. F^+ a^-

b. F^- a^-

c. F^- a^+

d. F^+ a^+

9. Na Figura 5.17, quantos *crossovers* são necessários para produzir um exconjugante completamente prototrópico?

10. Na Figura 5.18 C, por que o *crossover* é demonstrado como ocorrendo nos segmentos em laranja do DNA?

11. Na Figura 5.19, quantas espécies bacterianas diferentes são demonstradas como tendo contribuído com DNA para o plasmídio pk214?

12. Na Figura 5.25, você consegue apontar algum fago da progênie que poderia transduzir?

13. Na Figura 5.28, quais são as características físicas das placas de fagos recombinantes?

14. Na Figura 5.29, você acredita que b^+ poderia ser transduzido em vez de a^+? Tão bem quanto a^+?

15. Na Figura 5.30, que genes demonstram as mais altas frequências de cotransdução?

16. Na Figura 5.32, o que os segmentos metade vermelhos, metade azuis representam?

17. Na Figura 5.33, qual é o mais raro genótipo de λ produzido no lisado inicial?

18. Na Figura 5.38, precisamente qual gene é finalmente identificado a partir

da sequência do genoma?

PROBLEMAS BÁSICOS

19. Descreva o estado do fator F em uma linhagem Hfr, F^+, e F^-.

20. Como uma cultura de células F^+ transfere marcadores do cromossomo hospedeiro para uma receptora?

21. Em relação à transferência gênica e à integração do gene transferido para o genoma receptor, compare:
a. Cruzamentos de Hfr por meio de conjugação e transdução generalizada.
b. Derivados de F', tais como F' *lac* e transdução especializada.

22. Por que a transdução generalizada é capaz de transferir qualquer gene, mas a transdução especializada é restrita a apenas um pequeno conjunto?

23. Uma geneticista microbiana isola uma nova mutação em *E. coli* e deseja mapear a sua localização cromossômica. Ela utiliza experimentos de cruzamento interrompido com linhagens Hfr e experimentos de transdução generalizada com o fago P1. Explique por que cada técnica, por si própria, é insuficiente para o mapeamento preciso.

24. Em *E. coli*, quatro linhagens Hfr doam os marcadores a seguir, demonstrados na ordem doada:

Linhagem 1:	M	Z	X	W	C
Linhagem 2:	L	A	N	C	W
Linhagem 3:	A	L	B	R	U
Linhagem 4:	Z	M	U	R	B

Todas essas linhagens Hfr são derivadas da mesma linhagem F^+. Qual é a ordem desses marcadores no cromossomo circular da F^+ original?

25. Você recebe duas linhagens de *E. coli*. A linhagem Hfr é arg^+ ala^+ glu^+ pro^+ leu^+ T^s; a linhagem F^- é arg^- ala^- glu^- pro^- leu^- T^r. Todos os marcadores são nutricionais, com exceção de *T*, que determina a sensibilidade ou a resistência ao fago T1. A ordem de entrada é conforme

fornecida, com *arg*$^+$ entrando na receptora primeiramente e *T*^{s} por último. Você observa que a linhagem F$^-$ morre quando exposta à penicilina (*pen*^{s}), mas a linhagem Hfr não morre (*pen*^{r}). Como você localizaria o *locus* de *pen* no cromossomo bacteriano em relação a *arg, ala, glu, pro* e *leu*? Formule a sua resposta em etapas lógicas e bem explicadas, e desenhe diagramas explícitos, quando possível.

26. É realizado um cruzamento entre duas linhagens de *E. coli*: Hfr *arg*$^+$ *bio*$^+$ *leu*$^+$ × F$^-$ *arg*$^-$ *bio*$^-$ *leu*$^-$. Estudos de cruzamento interrompido demonstram que *arg*$^+$ entra na receptora por último e, assim, recombinantes *arg*$^+$ são selecionados em um meio que contém apenas *bio* e *leu*. Esses recombinantes são testados em relação à presença de *bio*$^+$ e *leu*$^+$. São observados os números de indivíduos a seguir em relação a cada genótipo:

arg$^+$ *bio*$^+$ *leu*$^+$	320	*arg*$^+$ *bio*$^-$ *leu*$^+$	0
arg$^+$ *bio*$^+$ *leu*$^-$	8	*arg*$^+$ *bio*$^-$ *leu*$^-$	48

a. Qual é a ordem dos genes?

b. Quais são as distâncias de mapa em porcentagens de recombinação?

27. Os mapas de ligação em uma linhagem bacteriana Hfr são calculados em unidades de minutos (o número de minutos entre os genes indica o período de tempo necessário para que o segundo gene siga o primeiro na conjugação). Ao produzir os referidos mapas, geneticistas microbianos presumem que o cromossomo bacteriano seja transferido de Hfr para F$^-$ a uma velocidade constante. Portanto, presume-se que dois genes separados por 10 minutos perto da extremidade da origem estejam à mesma distância física que dois genes separados por 10 minutos perto da extremidade de ligação de F$^-$. Sugira um experimento crítico para testar a validade dessa hipótese.

28. Uma linhagem Hfr em particular normalmente transmite o marcador *pro*$^+$ como o último na conjugação. Em um cruzamento dessa linhagem com uma linhagem F$^-$, alguns recombinantes *pro*$^+$ são recuperados inicialmente no processo de cruzamento. Quando essas células *pro*$^+$ são misturadas com

células F^-, a maioria das células F^- é convertida em células *pro*$^+$ que também carreiam o fator F. Explique esses resultados.

29. Linhagens F' em *E. coli* são derivadas de linhagens Hfr. Em alguns casos, essas linhagens F' demonstram uma alta taxa de integração ao cromossomo bacteriano de uma segunda linhagem. Além disso, o sítio de integração com frequência é o local ocupado pelo fator sexual na linhagem Hfr original (antes da produção de linhagens F'). Explique esses resultados.

30. Você possui duas linhagens de *E. coli*, F^- *str*^{s} *ala*$^-$ e Hfr *str*^{s} *ala*$^+$, nas quais o fator F é inserido próximo de *ala*$^+$. Planeje um teste de triagem para detectar linhagens que carreiam F' *ala*$^+$.

31. Cinco linhagens de Hfr, A a E, são derivadas de uma única linhagem F^+ de *E. coli*. O quadro a seguir demonstra os tempos de entrada dos primeiros cinco marcadores em uma linhagem F^- quando cada um é utilizado em um experimento de conjugação interrompida:

A	B	C	D	E
mal$^+$ (1)	*ade*$^+$ (13)	*pro*$^+$ (3)	*pro*$^+$ (10)	*his*$^+$ (7)
str^s (11)	*his*$^+$ (28)	*met*$^+$ (29)	*gal*$^+$ (16)	*gal*$^+$ (17)
ser$^+$ (16)	*gal*$^+$ (38)	*xyl*$^+$ (32)	*his*$^+$ (26)	*pro*$^+$ (23)
ade$^+$ (36)	*pro*$^+$ (44)	*mal*$^+$ (37)	*ade*$^+$ (41)	*met*$^+$ (49)
his$^+$ (51)	*met*$^+$ (70)	*str*^{s} (47)	*ser*$^+$ (61)	*xyl*$^+$ (52)

a. Desenhe um mapa da linhagem F^+, indicando as posições de todos os

genes e suas distâncias em minutos.

b. Demonstre o ponto de inserção e a orientação do plasmídio F em cada linhagem Hfr.

c. Na utilização de cada uma dessas linhagens Hfr, declare qual alelo você selecionaria para obter a mais alta proporção de exconjugantes Hfr.

32. Células de *Streptococcus pneumoniae* de genótipo *str*^{s} *mtl*$^{-}$ são transformadas pelo DNA doador de genótipo *str*^{r} *mtl*$^{+}$ e (em um experimento em separado) por uma mistura de dois DNA com genótipos *str*^{r} *mtl*$^{-}$ e *str*^{s} *mtl*$^{+}$. A tabela que acompanha demonstra os resultados.

DNA transformante	*Porcentagem de células transformadas em*		
	str^{r} *mtl*$^{-}$	*str*^{s} *mtl*$^{+}$	*str*^{r} *mtl*$^{+}$
str^{r} *mtl*$^{+}$	4,3	0,40	0,17
str^{r} *mtl*$^{-}$ + *str*^{a} *mtl*$^{+}$	2,8	0,85	0,0066

a. O que a primeira fileira da tabela informa a você? Por quê?

b. O que a segunda fileira da tabela informa a você? Por quê?

33. Relembre que, no Capítulo 4, consideramos a possibilidade de que um evento de *crossover* afete a probabilidade de outro *crossover*. No bacteriófago T4, o gene *a* está a 1,0 u.m. do gene *b*, que está a 0,2 u.m. do gene *c*. A ordem dos genes é *a*, *b*, *c*. Em um experimento de recombinação, você recupera cinco *crossovers* duplos entre *a* e *c* de uma progênie de 100.000. É correto concluir que a interferência é negativa? Explique a sua resposta.

34. Células de *E. coli* foram infectadas por duas linhagens de vírus T4. Uma linhagem é mínima (*m*), lise rápida (*r*) e turva (*t*); a outra é do tipo selvagem em relação a todos os três marcadores. Os produtos líticos dessa

infecção foram plaqueados e classificados. As 10.342 placas resultantes foram distribuídas entre oito genótipos, como segue:

m r t	3.469	*m* + +	521
+ + +	3.727	+ *r t*	475
m r +	854	+ *r* +	171
m + *t*	163	+ + *t*	963

a. Quais são as distâncias de ligação entre *m* e *r*, entre *r* e *t*, e entre *m* e *t*.

b. Determine a ordem de ligação dos três genes.

c. Qual é o coeficiente de coincidência (ver Capítulo 4) nesse cruzamento? O que ele significa?

35. Com a utilização de P22 como um fago de transdução generalizada cultivado em uma doadora bacteriana pur^+ pro^+ his^+, uma linhagem receptora de genótipo pur^- pro^- his^- foi infectada e incubada. Posteriormente, transdutantes para pur^+, pro^+ e his^+ foram selecionados individualmente nos experimentos I, II e III, respectivamente.

a. Qual meio é utilizado em cada um desses experimentos de seleção?

b. Os transdutantes foram examinados em relação à presença de marcadores doadores não selecionados, com os resultados a seguir:

I	II	III
pro^- his^- 86%	pur^- his^- 44%	pur^- pro^- 20%
pro^+ his^- 0%	pur^+ his^- 0%	pur^+ pro^- 14%
pro^- his^+ 10%	pur^- his^+ 54%	pur^- pro^+ 61%
pro^+ his^+ 4%	pur^- his^+ 2%	pur^+ pro^+ 5%

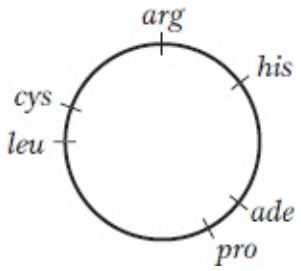
Qual é a ordem dos genes bacterianos?

c. Quais dois genes estão mais próximos?

d. Com base na sua resposta para a parte *c*, explique as proporções relativas dos genótipos observados no experimento II.

36. Embora a maior parte dos transdutantes *gal*$^+$ mediados por λ seja de lisógenos induzíveis, uma pequena porcentagem desses transdutantes de fato não é lisogênica (ou seja, eles não contêm λ integrado). Experimentos de controle demonstram que esses transdutantes não são produzidos por meio de mutação. Qual é a provável origem desses tipos?

37. Uma linhagem bacteriana *ade*$^+$ *arg*$^+$ *cys*$^+$ *his*$^+$ *leu*$^+$ *pro*$^+$ sabidamente é lisogênica em relação a um fago recentemente descoberto, mas o sítio do profago não é conhecido. O mapa bacteriano é



A linhagem lisogênica é utilizada como uma fonte do fago, e os fagos são adicionados a uma linhagem bacteriana de genótipo *ade*$^-$ *arg*$^-$ *cys*$^-$ *his*$^-$ *leu*$^-$ *pro*$^-$. Após uma breve incubação, amostras dessas bactérias são plaqueadas em seis meios diferentes, com as suplementações indicadas na tabela a seguir. A tabela também demonstra se foram observadas colônias nos diversos meios.

Meio	*Suplementação de nutrientes no meio*						Presença de colônias
	Ade	Arg	Cys	His	Leu	Pro	
1	–	+	+	+	+	+	N

2	+	–	+	+	+	+	N
3	+	+	–	+	+	+	C
4	+	+	+	–	+	+	N
5	+	+	+	+	–	+	C
6	+	+	+	+	+	–	N

(Nessa tabela, um sinal de mais indica a presença de um suplemento de nutriente, um sinal de menos indica que um suplemento não está presente, N indica ausência de colônias, e C indica presença de colônias.)

a. Qual processo genético está atuando aqui?

b. Qual é o *locus* aproximado do profago?

38. Em um sistema de transdução generalizada com a utilização do fago P1, a doadora é pur^+ nad^+ pdx^- e a receptora é pur^- nad^- pdx^+. O alelo doador pur^+ é selecionado inicialmente após a transdução e 50 transdutantes pur^+ são pontuados em seguida em relação aos outros alelos presentes. Aqui estão os resultados:

Genótipo	Número de colônias
nad^+ pdx^+	3
nad^+ pdx^-	10
nad^- pdx^+	24
nad^- pdx^-	13

a. Qual é a frequência de cotransdução em relação a *pur* e *nad*?

b. Qual é a frequência de cotransdução em relação a *pur* e *pdx*?

c. Quais dos *loci* não selecionados estão mais próximos de *pur*?

d. *nad* e *pdx* estão do mesmo lado ou em lados opostos de *pur*? Explique.

(Desenhe as trocas necessárias para produzir as diversas classes transformantes sob cada ordem para verificar qual requer o número mínimo para produzir os resultados obtidos.)

39. Em um experimento de transdução generalizada, fagos são coletados de uma linhagem doadora de *E. coli* de genótipo cys^+ leu^+ thr^+ e utilizados para transduzir uma receptora de genótipo cys^- leu^- thr^-. Inicialmente, a população receptora tratada é plaqueada em um meio mínimo suplementado com leucina e treonina. Muitas colônias são obtidas.

a. Quais são os possíveis genótipos dessas colônias?

b. Essas colônias em seguida são plaqueadas em réplica em três meios diferentes: (1) mínimo mais treonina apenas; (2) mínimo mais leucina apenas; e (3) mínimo. Quais genótipos poderiam, teoricamente, crescer nesses três meios?

c. Das colônias originais, observa-se que 56% crescem no meio 1, 5% no meio 2 e nenhuma colônia no meio 3. Quais são os genótipos reais das colônias nos meios 1, 2 e 3?

d. Desenhe um mapa demonstrando a ordem dos três genes e qual dos dois genes externos está mais próximo do gene intermediário.

40. Deduza os genótipos das linhagens 1 a 4 de *E. coli* a seguir:

41. Em um experimento de conjugação interrompida em *E. coli*, o gene *pro* entra após o gene *thi*. Uma linhagem Hfr pro^+ thi^+ é cruzada com uma linhagem F^- pro^- thi^- e exconjugantes são plaqueados em meio que contém tiamina, mas não prolina. É observado um total de 360 colônias e elas são isoladas e cultivadas em meio totalmente suplementado. Em seguida essas culturas são testadas em relação à sua capacidade de crescer em meio que não contém prolina ou tiamina (meio mínimo) e observa-se que 320 colônias conseguem crescer, mas o restante não.

a. Deduza os genótipos dos dois tipos de culturas.

b. Desenhe os eventos de *crossover* necessários para produzir esses genótipos.

c. Calcule a distância entre os genes *pro* e *thi* em unidades de recombinação.

Como solucionar o problema 41

1. Que tipo de organismo é a *E. coli*?
2. Qual é a aparência de uma cultura de *E. coli*?
3. Em quais tipos de substratos a *E. coli* em geral cresce em seu *habitat* natural?
4. Quais são as exigências mínimas para que as células de *E. coli* se dividam?

5. Defina os termos *prototrófico* e *auxotrófico*.
6. Quais culturas nesse experimento são prototróficas e quais são auxotróficas?
7. Tendo recebido algumas linhagens de genótipo desconhecido quanto a tiamina e prolina, como você testaria os seus genótipos? Forneça os detalhes experimentais precisos, incluindo o equipamento.
8. Quais tipos de substâncias químicas são a prolina e a tiamina? Isso é importante nesse experimento?
9. Desenhe um diagrama demonstrando o conjunto integral de manipulações realizadas no experimento.
10. Por que você acredita que o experimento tenha sido realizado?
11. Como foi estabelecido que *pro* entra depois de *thi*? Forneça as etapas experimentais precisas.
12. De que modo um experimento de cruzamento interrompido difere do experimento descrito nesse problema?
13. O que é um exconjugante? Como você acredita que os exconjugantes foram obtidos? (Isso pode incluir genes não descritos nesse problema.)
14. Quando se diz que o gene *pro* entra depois de *thi*, isso significa o alelo *pro*, o alelo pro^+, qualquer um deles ou ambos?
15. O que é "meio totalmente suplementado" no contexto dessa questão?
16. Alguns exconjugantes não crescem em meio mínimo. Em qual meio eles crescem?
17. Declare os tipos de *crossovers* que participam na recombinação Hfr × F^-. Como esses *crossovers* diferem dos *crossovers* em eucariotos?
18. O que é uma unidade de recombinação no contexto da presente análise? Como ela difere das unidades de mapa utilizadas na genética de eucariotos?

42. Um experimento de transdução generalizada utiliza uma linhagem $metE^+$ $pyrD^+$ como doadora e $metE^-$ $pyrD^-$ como receptora. Os transdutantes $metE^+$ são selecionados e em seguida testados em relação ao alelo $pyrD^+$.

Foram obtidos os números a seguir:

$metE^+$ $pyrD^-$	857
$metE^+$ $pyrD^+$	1

Esses resultados sugerem que esses *loci* estão ligados de modo próximo? Quais outras explicações existem em relação ao único "duplo"?

43. Uma linhagem $argC^-$ foi infectada com um fago transdutor e o lisado foi utilizado para transduzir receptoras $metF^-$ em um meio que contém arginina, mas não metionina. As transdutantes $metF^+$ em seguida foram testadas em relação à necessidade de arginina: a maior parte era $argC^+$, mas observou-se que uma pequena porcentagem era $argC^-$. Desenhe diagramas para demonstrar a provável origem das linhagens $argC^+$ e $argC^-$.

PROBLEMAS DESAFIADORES

44. Quatro linhagens de *E. coli* de genótipo a^+ b^- são rotuladas 1, 2, 3 e 4. Quatro linhagenss de genótipo a^- b^+ são rotuladas 5, 6, 7 e 8. Os dois genótipos são misturados em todas as combinações possíveis e (após a incubação) são plaqueadas para determinar a frequência de recombinantes a^+ b^+. São obtidos os resultados a seguir, em que M = Muitos recombinantes, L = Poucos recombinantes e 0 = Nenhum recombinante.

	1	2	3	4
5	0	M	M	0
6	0	M	M	0
7	L	0	0	M

8	0	L	L	0

Com base nesses resultados, atribua um tipo de sexo (seja Hfr, F^+ ou F^-) a cada linhagem.

45. Uma linhagem Hfr de genótipo $a^+ b^+ c^+ d^- str^s$ é cruzada com uma linhagem fêmea de genótipo $a^- b^- c^- d^+ str^r$. Em diversos momentos, os pares do cruzamento são separados por meio da agitação vigorosa da cultura. Em seguida as células são plaqueadas em três tipos de ágar, conforme demonstrado a seguir, onde o nutriente A possibilita o crescimento de células a^-; o nutriente B, de células b^-; o nutriente C, de células c^-; e o nutriente D, de células d^-. (Um sinal de mais indica a presença de estreptomicina ou de um nutriente, e um sinal de menos indica a sua ausência.)

Tipo de ágar	Str	A	B	C	D
1	+	+	+	–	+
2	+	–	+	+	+
3	+	+	–	+	+

a. Quais genes doadores estão sendo selecionados em cada tipo de ágar?

b. A tabela a seguir demonstra o número de colônias em cada tipo de ágar em relação a amostras coletadas em diversos tempos após a mistura das linhagens. Utilize essas informações para determinar a ordem dos genes *a*, *b* e *c*.

Tempo da amostragem (min)	*Número de colônias em ágar do tipo*		
	1	2	3

0	0	0	0
5	0	0	0
7,5	102	0	0
10	202	0	0
12,5	301	0	74
15	400	0	151
17,5	404	49	225
20	401	101	253
25	398	103	252

c. De cada uma das placas aos 25 minutos, 100 colônias são coletadas e transferidas para uma placa de Petri contendo ágar com todos os nutrientes, exceto D. Os números de colônias que crescem nesse meio são: 90 em relação à amostra do ágar tipo 1, 52 em relação à amostra do ágar tipo 2, e 9 em relação à amostra do ágar tipo 3. Com a utilização desses dados, encaixe o gene *d* na sequência de *a, b* e *c*.

d. Em que tempo da amostragem você espera que as colônias apareçam pela primeira vez no ágar que contém C e estreptomicina, mas não A ou B?

46. No cruzamento Hfr *aro*$^{+}$ *arg*$^{+}$ *ery*^{r} *str*^{s} × F^{-} *aro*$^{-}$ *arg*$^{-}$ *ery*^{s} *str*^{r}, os marcadores são transferidos na ordem fornecida (com *aro*$^{+}$ entrando primeiramente), mas os primeiros três genes estão muito próximos. Exconjugantes são plaqueados em um meio contendo Str (estreptomicina, para matar as células Hfr), Ery (eritromicina), Arg (arginina), e Aro (aminoácidos aromáticos). São obtidos os resultados a seguir em relação a

300 colônias isoladas a partir dessas placas e testadas em relação ao crescimento em diversos meios: em Ery apenas, 263 linhagens cresceram; em Ery + Arg, 264 linhagens cresceram; em Ery + Aro, 290 linhagens cresceram; em Ery + Arg + Aro, 300 linhagens cresceram.

a. Desenhe uma lista de genótipos, e indique o número de indivíduos em cada genótipo.

b. Calcule as frequências de recombinação.

c. Calcule a proporção do tamanho da região de *arg* a *aro* e do tamanho da região de *ery* a *arg*.

47. É realizada uma transformação bacteriana com uma linhagem doadora que é resistente a quatro fármacos, A, B, C e D, e uma linhagem receptora que é sensível a todos os quatro fármacos. A população resultante de células receptoras é dividida e plaqueada em meios que contêm diversas combinações dos fármacos. A tabela a seguir demonstra os resultados:

Fármacos adicionados	Número de colônias	Fármacos adicionados	Número de colônias
Nenhum	10.000	BC	50
A	1.155	BD	48
B	1.147	CD	785
C	1.162	ABC	31
D	1.140	ABD	43
AB	47	ACD	631
AC	641	BCD	35

a. Um dos genes está distante dos outros três, que aparentam estar ligados de modo próximo. Qual é o gene distante?

b. Qual é a ordem provável dos três genes ligados de modo próximo?

48. Você possui duas linhagens de λ que conseguem lisogenizar *E. coli*; seus mapas de ligação são como segue:

O segmento demonstrado na parte inferior do cromossomo, designado 1-2-3, é a região responsável pelo pareamento e pelo *crossover* com o cromossomo de *E. coli*. (Mantenha os marcadores em todos os seus desenhos.)

a. Diagrame o modo pelo qual λ da linhagem X é inserido no cromossomo de *E. coli* (de modo que *E. coli* seja lisogenizada).

b. As bactérias que são lisogênicas em relação à linhagem X podem ser superinfectadas por meio da utilização da linhagem Y. Uma determinada porcentagem dessas bactérias superinfectadas se torna "duplamente" lisogênica (ou seja, lisogênica em relação a ambas as linhagens). Diagrame como isso ocorrerá. (Não se preocupe a respeito de como as duplamente lisogênicas são detectadas.)

c. Diagrame como os dois profagos λ conseguem parear.

d. Podem ser recuperados os produtos de *crossover* entre os dois profagos. Diagrame um evento de *crossover* e as consequências.

49. Você possui três linhagens de *E. coli*: a linhagem A é F' cys^+ *trp1*/cys^+ *trp1* (ou seja, tanto F' quanto o cromossomo carreiam cys^+ e *trp1*, um alelo em

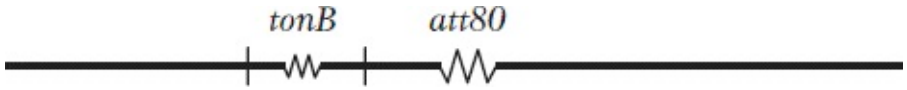
relação à necessidade de triptofano). A linhagem B é F^{-} *cys*$^{-}$ *trp2 Z* (essa linhagem necessita de cisteína para crescer e carreia *trp2*, outro alelo que causa necessidade de triptofano; a linhagem B é lisogênica em relação ao fago Z de transdução generalizada). A linhagem C é F^{-} *cys*$^{+}$ *trp1* (ela é uma F-derivada da linhagem A que perdeu o F'). Como você determinaria se *trp1* e *trp2* são alelos do mesmo *locus*? (Descreva os cruzamentos e os resultados esperados.)

50. Um fago de transdução generalizada é utilizado para transduzir uma linhagem receptora a^{-} b^{-} c^{-} d^{-} e^{-} de *E. coli* com uma doadora a^{+} b^{+} c^{+} d^{+} e^{+}. A cultura receptora é plaqueada em diversos meios, com os resultados demonstrados na tabela a seguir. (Observe que a^{-} indica uma necessidade de A como um nutriente, e assim por diante.) O que você pode concluir a respeito da ligação e da ordem dos genes?

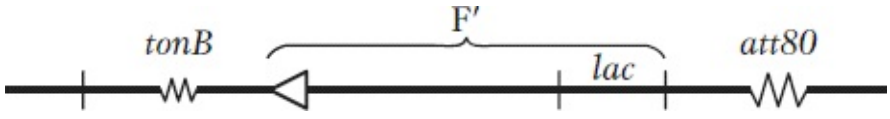
Compostos adicionados ao meio mínimo	Presença (+) ou ausência (—) de colônias
CDE	–
BDE	–
BCE	+
BCD	+
ADE	–
ACE	–
ACD	–
ABE	–

ABD	+
ABC	–

51. Em 1965, Jon Beckwith e Ethan Signer elaboraram um método para a obtenção de fagos transdutores especializados que carreiam a região *lac*. Eles sabiam que o sítio de integração, designado *att80*, em relação ao fago temperado ϕ80 (um parente do fago λ) estava localizado perto de *tonB*, um gene que confere resistência ao fago virulento T1:



Eles utilizaram um plasmídio F' *lac*$^+$ que não podia ser replicado em altas temperaturas em uma linhagem que carreava uma deleção dos genes *lac*. Ao forçar a célula a permanecer *lac*$^+$ a altas temperaturas, os pesquisadores puderam selecionar linhagens nas quais o plasmídio integrou-se ao cromossomo, possibilitando, assim que F' *lac* fosse conservado em altas temperaturas. Ao combinar essa seleção com uma seleção simultânea em relação à resistência à infecção pelo fago T1, eles observaram que as únicas sobreviventes eram células nas quais o F' *lac* integou-se ao *locus tonB*, conforme demonstrado aqui:



Esse resultado posicionou a região *lac* perto do sítio de integração para o fago ϕ80. Descreva as etapas subsequentes que os pesquisadores necessariamente seguiram para isolar as partículas de transdução especializadas do fago ϕ80 que carreavam a região *lac*.

52. *E. coli* do tipo selvagem capta e concentra um determinado corante alimentício vermelho, tornando as colônias vermelho-sangue. Foi utilizada a mutagênese por transpóson, e as células foram plaqueadas em corante

alimentício. A maior parte das colônias era vermelha, mas algumas colônias não captaram o corante e apareceram brancas. Em uma colônia branca, o DNA adjacente ao transpóson inserido foi sequenciado, com a utilização de um *primer* de replicação de DNA idêntico à parte terminal da sequência do transpóson, e observou-se que a sequência adjacente ao transpóson corresponde a um gene de função desconhecida denominado *atoE*, abrangendo as posições 2,322 a 2,324 Mb no mapa (numerada a partir de uma posição arbitrária zero). Proponha uma função para *atoE*. Qual processo biológico poderia ser investigado desse modo, e quais outros tipos de colônias brancas poderiam ser esperados?

CAPÍTULO 6

Interação Gênica

As cores dos pimentões são determinadas pela interação de diversos genes. Um alelo *Y* promove a eliminação precoce da clorofila (um pigmento verde), enquanto *y* não promove. O alelo *R* determina o pigmento carotenoide vermelho e *r* determina o amarelo. Os alelos *c1* e *c2* de dois genes diferentes infrarregulam as quantidades de carotenoides, causando as tonalidades mais claras. Laranja é vermelho infrarregulado. Marrom é verde mais vermelho. Amarelo-pálido é produto da infrarregulação de amarelo. (Anthony Griffiths.)

TÓPICOS

6.1 Interações de alelos de um único gene | Variações de dominância
6.2 Interação dos genes nas vias
6.3 Inferência das interações gênicas
6.4 Penetrância e expressividade

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Desenhar experimentos para testar duas ou mais mutações em relação ao alelismo, com a utilização das proporções da progênie ou com a utilização de testes de complementação
- Inferir vários tipos de dominância com base nos fenótipos dos heterozigotos
- Reconhecer o diagnóstico em relação à presença de um alelo letal
- Inferir a interação de diferentes genes, com base nas proporções mendelianas modificadas
- Formular hipóteses moleculares razoáveis para explicar vários tipos de interação gênica
- Reconhecer os diagnósticos em relação a variações na penetrância e na expressividade dos genótipos
- Prever a progênie de cruzamentos nos quais os genes demonstram um ou mais dos tipos anteriores de interação.

A motivação da nossa apresentação no livro até agora tem sido demonstrar como os geneticistas identificam um gene que afeta alguma propriedade biológica de interesse. Observamos como as abordagens da genética direta podem ser utilizadas para identificar genes individuais. O pesquisador inicia com um conjunto de mutantes e em seguida cruza cada um dos mutantes com o tipo selvagem para verificar se o mutante demonstra herança monogênica. Os dados cumulativos de um referido programa de pesquisa revelariam um conjunto

de genes no qual todos apresentam papéis no desenvolvimento da propriedade em investigação. Em alguns casos, o pesquisador pode ser capaz de identificar funções bioquímicas específicas em relação a muitos dos genes por meio da comparação das sequências gênicas com aquelas de outros organismos. A próxima etapa, que é um desafio maior, é deduzir como os genes em conjunto interagem para influenciar o fenótipo.

Como as interações gênicas subjacentes a uma propriedade são deduzidas? Uma abordagem molecular é analisar as interações proteicas diretamente *in vitro* por meio da utilização de uma proteína como "isca" e observar quais outras proteínas celulares se unem a ela. As proteínas que se observa se ligarem à isca são candidatas à interação na célula viva. Outra abordagem molecular é analisar transcritos de mRNA. Os genes que colaboram em algum processo específico do desenvolvimento podem ser definidos por meio do conjunto de RNA transcritos, presentes quando aquele processo está ocorrendo, um tipo de análise atualmente realizada com a utilização de *chips* genômicos (ver Capítulo 14). Finalmente, as interações gênicas e sua significância na modelagem do fenótipo podem ser deduzidas por meio da *análise genética*, que é o foco deste capítulo.

As interações gênicas podem ser amplamente classificadas em duas categorias. A primeira categoria consiste em interações dos alelos de um *locus*, falando amplamente, variações de dominância. Nos capítulos anteriores, lidamos com a dominância total e a recessividade total, mas conforme veremos neste capítulo, existem outros tipos de dominância, cada uma com a sua própria biologia celular subjacente. Embora essa informação não aborde a variedade de genes que afetam uma função, pode-se aprender muito sobre o papel de um gene ao considerar as interações alélicas. A segunda categoria consiste em interações de dois ou mais *loci*. Essas interações revelam o número e os tipos de genes no programa geral subjacente a uma função biológica em particular.

6.1 Interações de alelos de um único gene | Variações de dominância

Existem milhares de diferentes modos de alterar a sequência de um gene, cada um produzindo um alelo mutante, embora apenas alguns desses alelos mutantes apareçam em uma população real. Os alelos mutantes conhecidos de um gene e seu alelo do tipo selvagem são denominados **alelos múltiplos**, ou **série alélica**.

Um dos testes realizados de modo rotineiro sobre um novo alelo mutante é verificar se ele é dominante ou recessivo. Informações básicas sobre a dominância e a recessividade são úteis para trabalhar com a nova mutação e podem ser uma fonte de percepção sobre o modo como o gene funciona, conforme veremos nos exemplos. A dominância é uma manifestação de como *os alelos de um gene único* interagem em um heterozigoto. Em qualquer experimento, os alelos que interagem podem ser alelos do tipo selvagem e mutantes (+/*m*) ou dois alelos mutantes diferentes (m_1/m_2). Foram descobertos diversos tipos de dominância, cada um deles representando um tipo diferente de interação dos alelos.

Dominância completa e recessividade

O tipo mais simples de dominância é a **dominância total**, ou **completa**, que examinamos no Capítulo 2. Um alelo totalmente dominante será expresso no fenótipo quando apenas uma cópia estiver presente, como em um heterozigoto, enquanto o alelo alternativo será totalmente recessivo. Na dominância completa, o homozigoto dominante não pode ser distinguido do heterozigoto; ou seja, no nível fenotípico, *A/A* = *A/a*. Conforme mencionado anteriormente, a fenilcetonúria (PKU) e muitas outras doenças humanas monogênicas são totalmente recessivas, enquanto seus alelos do tipo selvagem são dominantes. Outras doenças monogênicas, tais como a acondroplasia, são totalmente dominantes, enquanto, nesses casos, o alelo do tipo selvagem é recessivo. Como essas relações de dominância podem ser interpretadas no nível celular?

A doença PKU é um bom modelo geral em relação às mutações recessivas. Relembre que a PKU é causada por um alelo defeituoso do gene que codifica a enzima fenilalanina hidroxilase (PAH). Na ausência da PAH normal, a fenilalanina que entra no corpo pelo alimento não é fragmentada e, portanto, acumula-se. Nessas condições, a fenilalanina é convertida em ácido fenilpirúvico, que é transportado até o cérebro por meio da corrente sanguínea e, ali, impede o

desenvolvimento normal, levando ao retardo mental. O motivo de o alelo defeituoso ser recessivo é que uma "dose" do alelo *P* do tipo selvagem produz PAH suficiente para fragmentar a fenilalanina que entra no corpo. Portanto, diz-se que o alelo do tipo selvagem PAH é *haplossuficiente*. *Haplo* significa uma dose haploide e *suficiente* faz referência à capacidade de aquela dose única produzir o fenótipo do tipo selvagem. Portanto, ambos *P/P* (duas doses) e *P/p* (uma dose) apresentam atividade de PAH suficiente para resultar na bioquímica celular normal. Pessoas com *p/p* apresentam zero dose de atividade de PAH. A Figura 6.1 ilustra a noção geral.

Como podemos explicar as mutações totalmente dominantes? Existem diversos mecanismos moleculares para dominância. Um mecanismo encontrado regularmente é que o alelo do tipo selvagem de um gene é *haploinsuficiente*. Na haploinsuficiência, uma dose do tipo selvagem *não* é suficiente para alcançar níveis de função normais. Presuma que 16 unidades do produto de um gene sejam necessárias para a bioquímica normal e que cada alelo do tipo selvagem possa produzir 10 unidades. Dois alelos do tipo selvagem produzirão 20 unidades do produto, bem superiores ao mínimo. Mas considere o que ocorre se uma das mutações é uma **mutação nula**, que produz uma proteína não funcional. Uma mutação nula em combinação com um único alelo do tipo selvagem produziria 10 + 0 = 10 unidades, bem inferiores ao mínimo. Portanto, o heterozigoto (tipo selvagem/nulo) é mutante e a mutação é, por definição, dominante. Em camundongos, o gene *Tbx1* é haploinsuficiente. Esse gene codifica uma proteína reguladora de transcrição (um *fator de transcrição*) que atua nos genes responsáveis pelo desenvolvimento da faringe. Um nocaute de um dos alelos do tipo selvagem resulta em uma concentração inadequada da proteína reguladora, que resulta em defeitos no desenvolvimento das artérias faríngeas. Acredita-se que a mesma haploinsuficiência seja responsável pela síndrome de DiGeorge em seres humanos, uma condição com anormalidades cardiovasculares e craniofaciais.

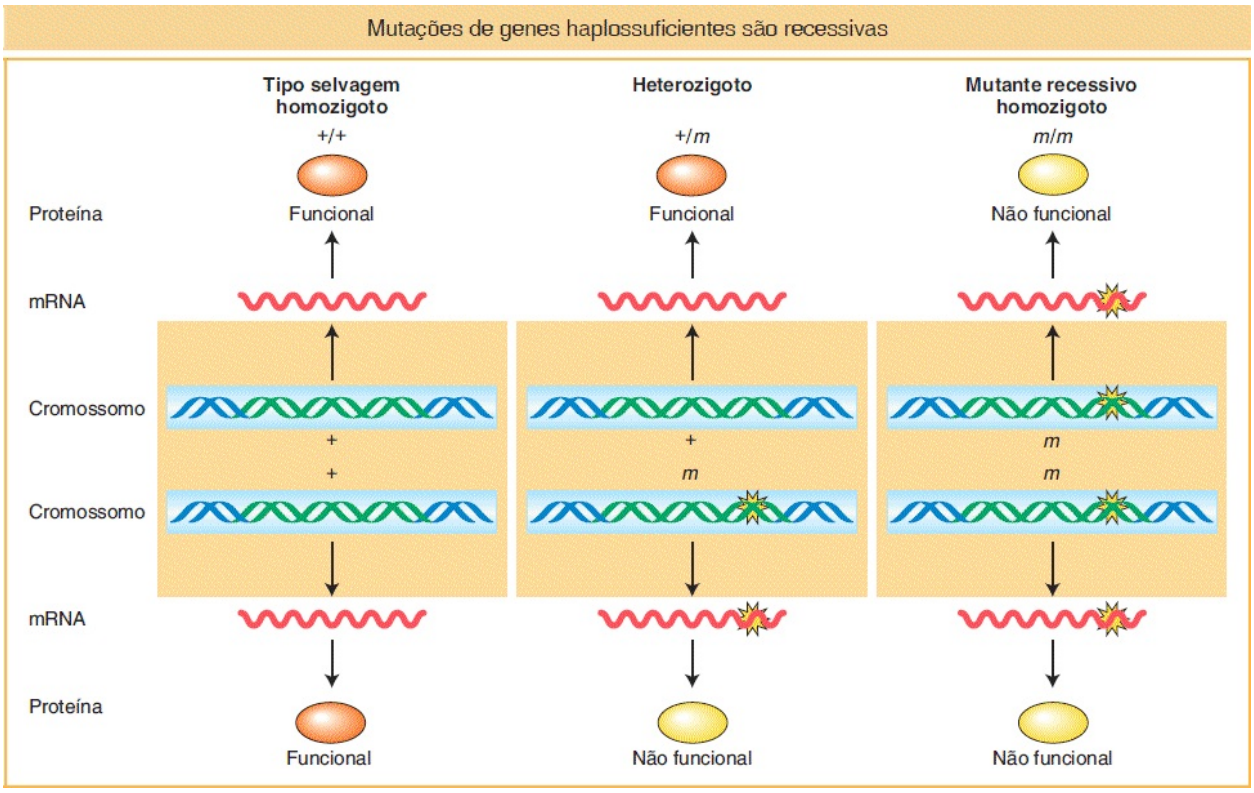


FIGURA 6.1 No heterozigoto, embora a cópia mutada do gene produza uma proteína não funcional, a cópia do tipo selvagem gera proteína funcional suficiente para produzir o fenótipo do tipo selvagem.

Outro tipo importante de mutação dominante é denominado **dominante negativo**. Polipeptídios com esse tipo de mutação atuam como "espoliadores" ou "trapaceiros". Em alguns casos, o produto do gene é uma unidade de uma proteína *homodimérica*, uma proteína composta por duas unidades do mesmo tipo. No heterozigoto (+/*M*), o polipeptídio mutante se liga ao polipeptídio do tipo selvagem e atua como um espoliador ao distorcê-lo ou ao interferir de outro modo em sua função. O mesmo tipo de estrago também pode impedir o funcionamento de um *heterodímero* composto por polipeptídios de diferentes genes. Em outros casos, o produto gênico é um monômero e, nessas situações, o mutante se liga ao substrato e atua como um espoliador ao dificultar a ligação da proteína do tipo selvagem ao substrato.

Um exemplo de mutações que podem atuar como dominantes negativas é observado no gene da proteína colágeno. Algumas mutações nesse gene dão origem ao fenótipo humano osteogênese imperfeita. O colágeno é uma proteína do

tecido conjuntivo formada por três monômeros entrelaçados (um trímero). No heterozigoto mutante, a proteína anormal é enrola-se ao redor de uma ou duas normais e distorce o trímero, levando ao mau funcionamento. Desse modo, o colágeno defeituoso atua como um espoliador. A diferença entre a haploinsuficiência e a ação de um dominante negativo como causas da dominância de uma mutação está ilustrada na Figura 6.2.

> CONCEITO-CHAVE Em relação à maior parte dos genes, uma única cópia do tipo selvagem é adequada para a expressão total (os referidos genes são haplossuficientes) e suas mutações nulas são totalmente recessivas. As mutações prejudiciais de genes haploinsuficientes com frequência são dominantes. Mutações nos genes que codificam unidades em homodímeros ou heterodímeros podem se comportar como dominantes negativas, que atuam por meio de proteínas “espoliadoras”.

Dois modelos em relação à dominância de uma mutação

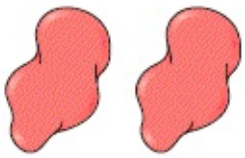
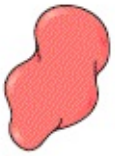
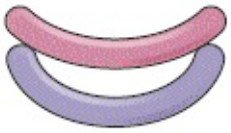
	Modelo 1: Haploinsuficiência	Modelo 2: Dominante negativo	Fenótipo
+/+	 2 "doses" do produto	Dímero	Tipo selvagem
M/*M*	0 "dose"		Mutante
+/*M*	 1 "dose" (inadequada)		Mutante

FIGURA 6.2 Uma mutação pode ser dominante em virtude (*esquerda*) de um único gene do tipo selvagem não produzir produto proteico suficiente para a função adequada ou (*direita*) de o alelo mutante atuar como um dominante negativo que produz um produto proteico "espoliador" (*spoiler*).

Dominância incompleta

Maravilhas são plantas nativas da América tropical cujas flores se abrem ao fim da tarde. Quando uma linhagem de maravilha do tipo selvagem pura que apresenta pétalas vermelhas é cruzada com uma linhagem pura que apresenta pétalas brancas, a F_1 apresenta pétalas rosa. Se uma F_2 for produzida por meio de autofecundação da F_1, o resultado é:

$\frac{1}{4}$ das plantas apresenta pétalas vermelhas

$\frac{1}{4}$ das plantas apresenta pétalas rosa

$\frac{1}{4}$ das plantas apresenta pétalas brancas

A Figura 6.3 demonstra esses fenótipos. A partir dessa proporção de 1:2:1 na F_2, podemos deduzir que o padrão de herança tem por base dois alelos de um gene único. Entretanto, os heterozigotos (a F_1 e metade da F_2) são de fenótipo intermediário. Ao inventar símbolos alélicos, podemos listar os genótipos das maravilhas nesse experimento como c^+/c^+ (vermelho), c/c (branco) e c^+/c (rosa). A ocorrência do fenótipo intermediário sugere uma **dominância incompleta**, o termo utilizado para descrever o caso geral no qual o fenótipo de um heterozigoto é intermediário entre aqueles dos dois homozigotos, em alguma escala de medição quantitativa.

Como explicamos a dominância incompleta no nível molecular? Na dominância incompleta, cada alelo do tipo selvagem em geral produz uma dose estabelecida de seu produto proteico. O número de doses de um alelo do tipo selvagem determina a concentração de uma substância química produzida pela proteína, tal como um pigmento. Na planta maravilha, duas doses produzem a maior parte das cópias do transcrito, produzindo, assim, a maior quantidade de proteína e, portanto, maior quantidade de pigmento, suficiente para tornar as pétalas das flores vermelhas. Uma dose produz menos pigmento e, assim, as pétalas são cor-de-rosa. Uma dose zero não produz pigmento.

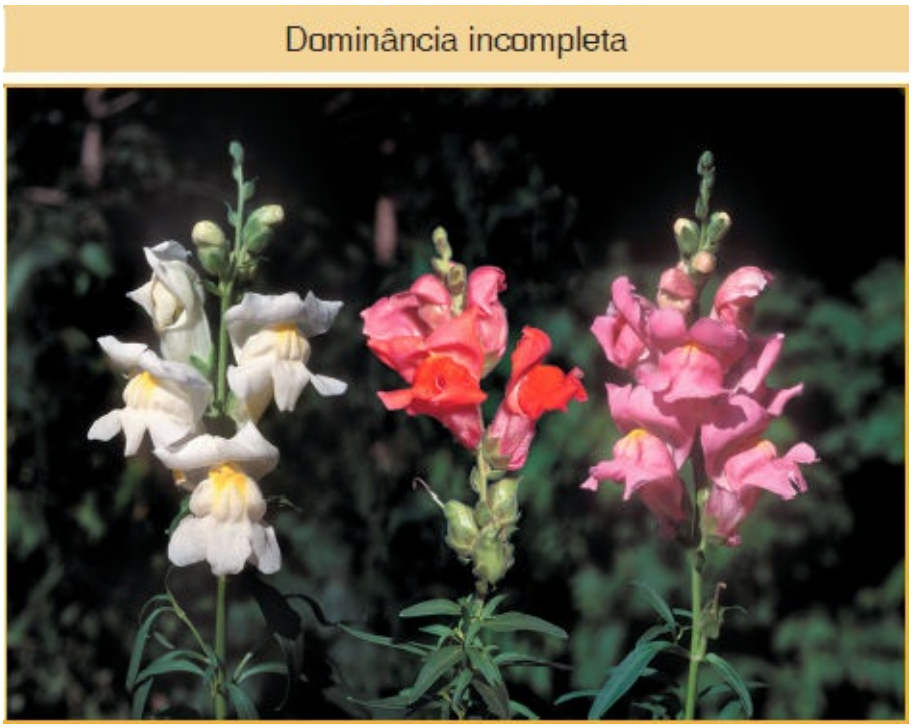


FIGURA 6.3 Em bocas-de-leão, um heterozigoto é rosa, o intermediário entre os dois homozigotos, o vermelho e o branco. O heterozigoto rosa demonstra dominância incompleta. (John Kaprielian/Science Source.)

Codominância

Outra variação sobre o tema da dominância é a **codominância**, a expressão de ambos os alelos de um heterozigoto. Um exemplo claro é observado nos grupos sanguíneos AB0 humanos, em que existe codominância de alelos de antígenos. Os grupos sanguíneos AB0 são determinados por três alelos de um gene. Esses três alelos interagem de diversos modos para produzir os quatro tipos sanguíneos do sistema AB0. Os três alelos principais são i, I^A e I^B, mas uma pessoa pode apresentar apenas dois dos três alelos ou duas cópias de um deles. As combinações resultam em seis genótipos diferentes: três homozigotos e três tipos diferentes de heterozigotos, como segue.

Genótipo	Tipo sanguíneo
I^A/I^A, I^A/i	A

I^B/I^B, I^B/i	B
I^A/I^B	AB
i/i	0

Nessa série alélica, os alelos determinam a presença e o tipo de uma molécula de açúcar complexa presente sobre a superfície dos eritrócitos. Essa molécula de açúcar é um antígeno, uma molécula de superfície celular que pode ser reconhecida pelo sistema imune. Os alelos I^A e I^B determinam dois tipos diferentes dessa molécula de superfície celular. Entretanto, o alelo i resulta na ausência de uma molécula de superfície celular desse tipo (é um alelo nulo). Nos genótipos I^A/i e I^B/i, os alelos I^A e I^B são totalmente dominantes em relação a i. Entretanto, no genótipo I^A/I^B, cada um dos alelos produz seu próprio tipo de molécula de superfície celular e, assim, os alelos A e B são codominantes.

A doença humana anemia falciforme ilustra os modos um pouco arbitrários pelos quais classificamos a dominância. O gene em questão codifica a molécula hemoglobina, que é responsável pelo transporte de oxigênio nos vasos sanguíneos e é o principal constituinte dos eritrócitos. Existem dois alelos principais, Hb^A e Hb^S e os três genótipos possíveis apresentam fenótipos diferentes, como segue:

Hb^A/Hb^A:	normal; os eritrócitos nunca se tornam falciformes
Hb^S/Hb^S:	anemia grave, com frequência fatal; a hemoglobina anormal faz com que os eritrócitos apresentem um formato falciforme
Hb^A/Hb^S:	nenhuma anemia; os eritrócitos se tornam falciformes apenas sob baixas concentrações de oxigênio

A Figura 6.4 demonstra uma micrografia eletrônica de hemácias, incluído algumas hemácias falciformes. Em relação à presença ou à ausência de anemia, o alelo Hb^A é dominante. No heterozigoto, um único alelo Hb^A produz hemoglobina funcional suficiente para prevenir a anemia. Em relação ao formato da hemácia, entretanto, existe dominância incompleta, conforme demonstrado pelo fato que, no

heterozigoto, muitas das células apresentam um formato discretamente falciforme. Finalmente, em relação à própria hemoglobina, existe codominância. Os alelos Hb^A e Hb^S codificam dois diferentes tipos de hemoglobina, que diferem em um único aminoácido e ambos os tipos são sintetizados no heterozigoto. Os tipos A e S da hemoglobina podem ser separados por meio de eletroforese, uma vez que apresentam cargas elétricas diferentes (Figura 6.5). Observamos que pessoas homozigotas Hb^A/Hb^A apresentam um tipo de hemoglobina (A) e pessoas anêmicas apresentam outro (tipo S), que se movimenta mais lentamente no campo elétrico. Os heterozigotos apresentam ambos os tipos, A e S. Em outras palavras, existe codominância no nível molecular. A fascinante genética de populações dos alelos Hb^A e Hb^S será considerada no Capítulo 20.

FIGURA 6.4 A hemácia falciforme é causada por uma mutação no gene da hemoglobina. (Eye of Science/Science Source.)

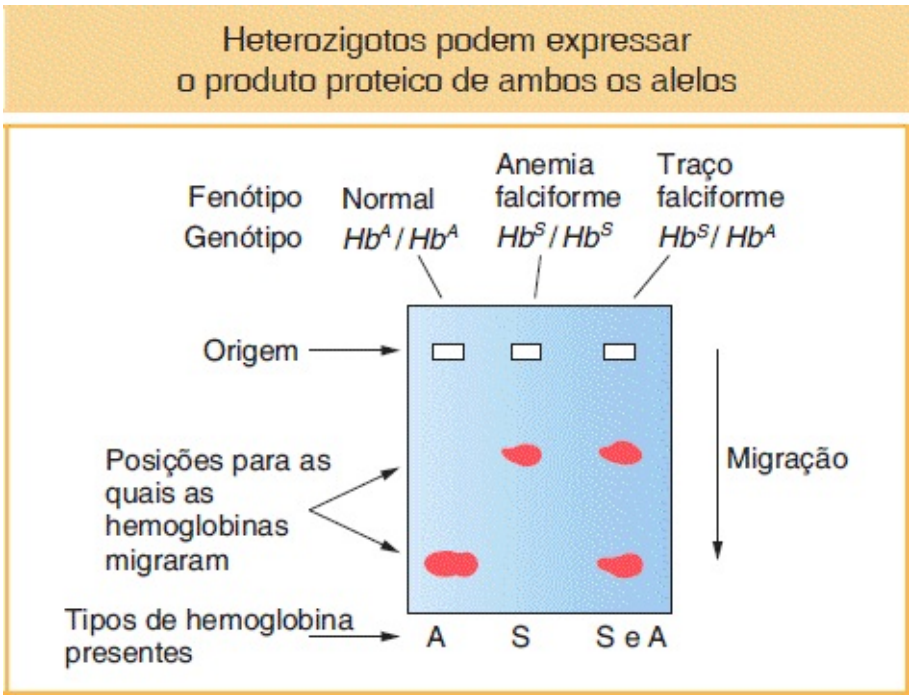


FIGURA 6.5 A eletroforese de hemoglobinas normal e mutante. Estão demonstrados os resultados produzidos pela hemoglobina de uma pessoa com traço falciforme (um heterozigoto), uma pessoa com anemia falciforme e uma pessoa normal. As manchas demonstram as posições até as quais as hemoglobinas migram no gel de amido.

A anemia falciforme ilustra a arbitrariedade dos termos *dominância, dominância incompleta* e *codominância*. O tipo de dominância inferido depende do nível fenotípico no qual o ensaio é realizado — do organismo, da célula ou da molécula. De fato, deve-se ter cautela com muitas das categorias que os cientistas utilizam para classificar as estruturas e os processos; essas categorias são planejadas por seres humanos para a conveniência da análise.

> CONCEITO-CHAVE Em geral, três tipos principais de dominância podem ser distinguidos: dominância total, dominância incompleta e codominância. O tipo de dominância é determinado pelas funções moleculares dos alelos de um gene e pelo nível investigativo da análise.

As folhas dos trevos demonstram diversas variações sobre o tema da

dominância. O trevo é o nome comum de plantas do gênero *Trifolium*. Existem muitas espécies. Algumas são nativas da América do Norte, enquanto outras crescem ali como sementes introduzidas. Foram realizadas muitas pesquisas genéticas com o trevo-branco, que demonstra variação considerável entre plantas individuais no curioso padrão em V, ou divisas, das folhas. Os diferentes tipos de divisas (e a ausência de divisas) são determinados por uma série de sete alelos, conforme observado na Figura 6.6, que demonstra os muitos tipos diferentes de possíveis interações em relação a até mesmo um alelo. Na maior parte dos casos práticos, muitos alelos de um gene podem ser observados juntos em uma população, constituindo uma série alélica. Os fenótipos demonstrados pelas combinações alélicas são muitos e variados, refletindo a natureza relativa da dominância: um alelo pode demonstrar dominância com um parceiro, mas não com outro. Portanto, a complexidade ilustrada pelo sistema do tipo sanguíneo AB0 é pequena em comparação àquela em um caso tal como as divisas do trevo.

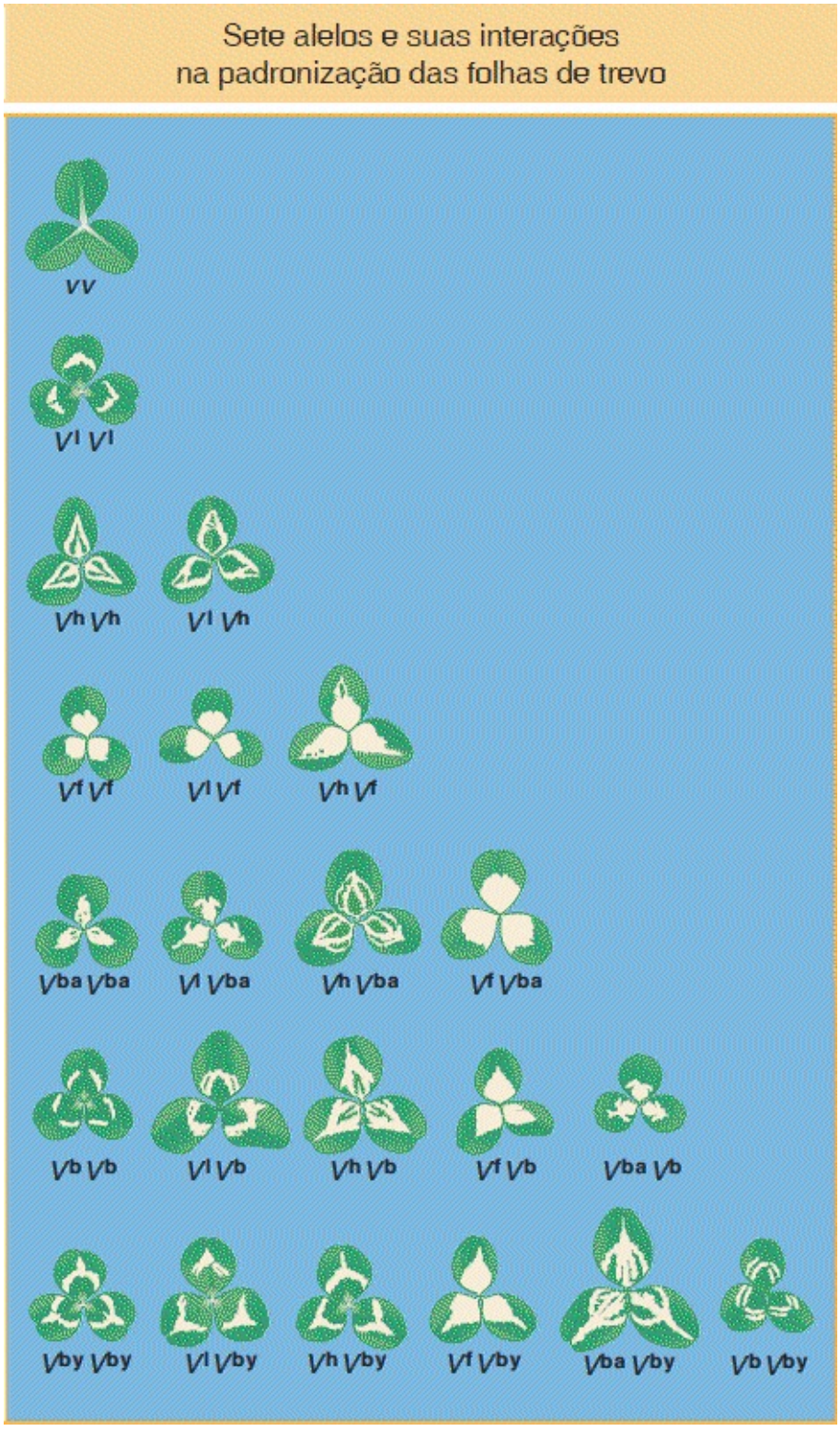


FIGURA 6.6 Alelos múltiplos determinam o padrão de divisas nas folhas do trevo-branco. O genótipo de cada planta está demonstrado abaixo dela. Existe uma variedade de interações de dominância. (Pesquisa por W. Ellis Davies.)

Alelos letais recessivos

Um alelo que é capaz de causar a morte de um organismo é denominado **alelo letal**. Na caracterização de um conjunto de alelos mutantes recentemente descobertos, por vezes observa-se que uma mutação recessiva é letal. Essa informação é potencialmente útil, no sentido em que demonstra que o gene recentemente descoberto (de função ainda desconhecida) é essencial para a operação do organismo. **Genes essenciais** são aqueles sem os quais um organismo morre. (Um exemplo de um gene essencial poderia ser um gene ribossômico, sem o qual nenhuma proteína seria produzida.) De fato, com a utilização da tecnologia do DNA moderna, um alelo mutante nulo de um gene de interesse pode agora ser produzido intencionalmente e tornado homozigoto para verificar se ele é letal e sob quais condições ambientais. Os alelos letais também são úteis na determinação do estágio de desenvolvimento no qual o gene normalmente atua. Nesse caso, os geneticistas observam se a morte em virtude de um alelo mutante letal ocorre inicial ou tardiamente no desenvolvimento de um zigoto. O fenótipo associado à morte também pode ser informativo em relação à função do gene; por exemplo, se um determinado órgão aparenta ser anormal, o gene provavelmente será expresso naquele órgão.

Qual é o teste diagnóstico em relação à letalidade? O teste é bem-ilustrado por um dos exemplos prototípicos de um alelo letal, um alelo de cor de pelagem em camundongos (ver Organismo-modelo, adiante). Camundongos do tipo selvagem normais apresentam pelagem com pigmentação geral um tanto escura. Uma mutação denominada *amarela* (uma cor de pelagem mais clara) demonstra um padrão de herança curioso. Se qualquer camundongo amarelo for cruzado com um camundongo do tipo selvagem homozigoto, sempre é observada uma proporção de 1:1 de camundongos amarelos e do tipo selvagem na progênie. Esse resultado sugere que um camundongo amarelo é sempre heterozigoto em relação ao alelo amarelo e que o alelo amarelo é dominante sobre o tipo selvagem. Entretanto, se dois camundongos amarelos forem cruzados entre si, o resultado é sempre como segue:

Amarelo × Amarelo → $\frac{2}{3}$ amarelos, $\frac{1}{3}$ do tipo selvagem

A Figura 6.7 demonstra uma ninhada típica a partir de um cruzamento entre camundongos amarelos.

Como a proporção de 2:1 pode ser explicada? Os resultados fazem sentido se for presumido que o alelo amarelo é letal quando homozigoto. Sabidamente o alelo amarelo é de um gene de cor de pelagem denominado *A*. Iremos denominá-lo A^{Y}. Portanto, os resultados do cruzamento de dois camundongos amarelos são:

FIGURA 6.7 Ninhada de um cruzamento entre dois camundongos heterozigotos em relação ao alelo dominante de cor de pelagem amarela. O alelo é letal em dose dupla. Nem toda a progênie é visível. (Anthony Griffiths.)

	$A^{Y}/A \times A^{Y}/A$	
Progênies	$\frac{1}{4}$ A^{Y}/A^{Y}	letal
	$\frac{1}{2}$ A^{Y}/A	amarelo
	$\frac{1}{4}$ A/A	tipo selvagem

A proporção mono-híbrida esperada de 1:2:1 seria observada entre os zigotos, mas é alterada para uma proporção de 2:1 na progênie realmente observada ao nascimento, tendo em vista que os zigotos com um genótipo A^{Y}/A^{Y} letal não sobrevivem para ser contados. Essa hipótese é amparada pela remoção dos úteros de fêmeas prenhes do cruzamento amarelo × amarelo; observa-se que um quarto

dos embriões está morto.

O alelo A^Y produz efeitos sobre duas características: cor da pelagem e sobrevivência. Entretanto, é totalmente possível que ambos os efeitos do alelo A^Y resultem da mesma causa básica, que promove pelagem amarela em dose única e morte em dose dupla. Em geral, o termo **pleiotrópico** é utilizado em relação a qualquer alelo que afete diversas propriedades de um organismo.

O fenótipo sem cauda em gatos Manx (Figura 6.8) também é produzido por um alelo que é letal no estado homozigoto. Uma dose única do alelo Manx, M^L, interfere substancialmente com o desenvolvimento da coluna vertebral, resultando na ausência de cauda no heterozigoto M^L/M. Mas no homozigoto M^L/M^L, a dose dupla do gene produz uma anormalidade tão extrema no desenvolvimento da coluna que o embrião não sobrevive.

Os alelos *amarelo* e M^L apresentam seus próprios fenótipos em um heterozigoto, mas a maioria dos letais recessivos é silenciosa no heterozigoto. Em uma referida situação, a letalidade recessiva é diagnosticada por meio da observação da morte de 25% da progênie em algum estágio do desenvolvimento.

ORGANISMO-MODELO *Camundongo*

O camundongo de laboratório é descendente do camundongo doméstico *Mus musculus*. As linhagens puras atualmente utilizadas como padrões são derivadas dos camundongos criados nos séculos anteriores por “apreciadores” de camundongos. Entre os organismos-modelo, é aquele cujo genoma se assemelha de modo mais próximo ao genoma humano. Seu número diploide de cromossomos é 40 (em comparação a 46 em seres humanos), e o genoma é discretamente menor do que aquele dos seres humanos (o genoma humano sendo de 3.000 Mb) e contém aproximadamente o mesmo número de genes (estimativa atual de 25.000). Além disso, todos os genes de camundongos aparentam ter correspondentes nos seres humanos. Uma grande proporção de genes está disposta em blocos exatamente nas mesmas posições daquelas dos seres humanos.

Pesquisas sobre a genética mendeliana de camundongos tiveram início no

começo do século 20. Uma das contribuições iniciais mais importantes foi a elucidação dos genes que controlam a cor e o padrão da pelagem. O controle genético da pelagem dos camundongos proporcionou um modelo para todos os mamíferos, incluindo gatos, cães, cavalos e gado. Também foi realizada uma grande quantidade de trabalhos sobre mutações induzidas por radiação e substâncias químicas. A genética de camundongos tem sido de grande significância na medicina. Uma grande proporção de doenças genéticas humanas apresenta, em camundongos, correspondentes úteis para estudos experimentais (denominados “modelos em camundongo”). O camundongo foi muito importante no desenvolvimento da nossa atual compreensão sobre o papel dos genes no câncer.

O genoma do camundongo pode ser modificado por meio da inserção de fragmentos específicos de DNA em um ovócito fertilizado ou em células somáticas. Os camundongos na fotografia receberam um gene de água-viva para a proteína fluorescente verde (GFP; do inglês, *green fluorescent protein*), que faz com que eles brilhem em verde sob luzes especiais. Também são possíveis nocautes e substituições de genes.

Uma limitação importante da genética de camundongos é o seu custo. Enquanto o trabalho com um milhão de *E. coli* ou *S. cerevisiae* é uma questão trivial, o trabalho com um milhão de camundongos requer uma edificação do tamanho de uma fábrica. Além disso, embora os camundongos reproduzam-se rapidamente (em comparação com os seres humanos), eles não conseguem competir com os microrganismos em função da velocidade do ciclo de vida dos microrganismos. Portanto, as seleções e as triagens em grande escala necessárias para detectar eventos genéticos raros não são possíveis.

Camundongos geneticamente modificados fluorescendo em verde. O gene de água-viva para a proteína fluorescente verde foi inserido nos cromossomos dos camundongos que fluorescem. Os outros camundongos são normais. (Eye of Science/Science Source.)

Ausência de cauda, um alelo letal recessivo em gatos

FIGURA 6.8 Um gato Manx. Um alelo dominante que causa a ausência da cauda é letal no estado homozigoto. O fenótipo de duas cores dos olhos não está relacionado com a ausência da cauda. (Gerard Lacz/NHPA/Photoshot.)

Se um alelo é letal ou não com frequência depende do ambiente no qual o organismo se desenvolve. Enquanto determinados alelos são letais em virtualmente qualquer ambiente, outros são viáveis em um ambiente, mas letais em outros. Doenças hereditárias humanas proporcionam alguns exemplos. A fibrose cística e a anemia falciforme são doenças que seriam letais sem tratamento. Além disso, muitos dos alelos favorecidos e selecionados por criadores de animais e cultivadores de plantas quase certamente seriam eliminados na natureza como um resultado da competição com os membros da população natural. As variedades mutantes de grão anão, que têm alta produção, fornecem bons exemplos; apenas o cultivo cuidadoso por parte dos fazendeiros manteve os referidos alelos para o nosso benefício.

Os geneticistas comumente encontram situações nas quais as proporções fenotípicas esperadas são consistentemente desviadas em uma direção em virtude

de um alelo mutante reduzir a viabilidade. Por exemplo, no cruzamento *A/a* × *a/a*, prevemos uma proporção de progênie de 50% *A/a* e 50% *a/a*, mas podemos observar consistentemente uma proporção tal como 55%:45% ou 60%:40%. Em um referido caso, diz-se que o alelo recessivo é *subletal*, tendo em vista que a letalidade é expressa em apenas alguns, mas não em todos, os indivíduos homozigotos. Portanto, a letalidade pode variar de 0 a 100%, dependendo do próprio gene, do restante do genoma e do ambiente.

Observamos que os alelos letais são úteis para diagnosticar o momento em que um gene atua e a natureza do defeito fenotípico que mata. Entretanto, a manutenção de estoques que contêm alelos letais para uso laboratorial é um desafio. Em diploides, os alelos letais recessivos podem ser mantidos como heterozigotos. Em haploides, alelos letais sensíveis ao calor são úteis. Eles são membros de uma classe geral de **mutações sensíveis à temperatura (st)**. Seu fenótipo é do tipo selvagem na **temperatura permissiva** (com frequência a temperatura ambiente), mas mutante em alguma **temperatura restritiva** mais alta. Acredita-se que os alelos sensíveis à temperatura sejam causados por mutações que tornam a proteína propensa à torção ou ao dobramento do seu formato até uma conformação inativa na temperatura restritiva. Estoques para pesquisas podem ser facilmente mantidos sob condições permissivas, e o fenótipo mutante pode ser analisado em um subconjunto de indivíduos por meio de uma alteração para as condições restritivas. As mutações letais dominantes sensíveis à temperatura também são úteis. Esse tipo de mutação é expressa até mesmo quando presente em uma dose única, mas apenas quando o experimentador altera o organismo para a temperatura restritiva.

Alelos nulos para genes identificados por sequenciamento genômico podem ser obtidos por meio da utilização de uma diversidade de procedimentos de “genética reversa” que nocauteiam especificamente a função daquele gene. Esses serão descritos no Capítulo 14.

> CONCEITO-CHAVE Para verificar se um gene é essencial, o alelo nulo é testado em relação à letalidade.

Agora nos voltamos para as abordagens que podem ser utilizadas para detectar a interação de dois ou mais *loci*.

6.2 Interação dos genes nas vias

Os genes atuam por meio do controle da química celular. No início do século 20, Archibald Garrod, um médico inglês (Figura 6.9), fez a primeira observação que ampara essa percepção. Garrod observou que diversas doenças humanas recessivas demonstram defeitos no que é denominado metabolismo, o conjunto geral de reações químicas que ocorrem em um organismo. Essa observação levou à noção de que as referidas doenças genéticas são “erros inatos do metabolismo”. Garrod trabalhou com uma doença denominada alcaptonúria (AKU), ou doença da urina negra. Ele descobriu que a substância responsável pela urina negra era o ácido homogentísico, que é encontrado em altas concentrações e que é secretado na urina em pacientes com AKU. Ele sabia que, em pessoas não afetadas, o ácido homogentísico é convertido em ácido maleilacetoacético; assim, ele propôs que, na AKU, existe um defeito nessa conversão. Consequentemente, o ácido homogentísico se acumula e é excretado. As observações de Garrod levantaram a possibilidade de que as vias químicas das células estivessem sob o controle de um grande conjunto de genes em interação. Entretanto, a demonstração direta desse controle foi fornecida pelo trabalho posterior de Beadle e Tatum com o fungo *Neurospora*.

Descobridor dos erros inatos do metabolismo

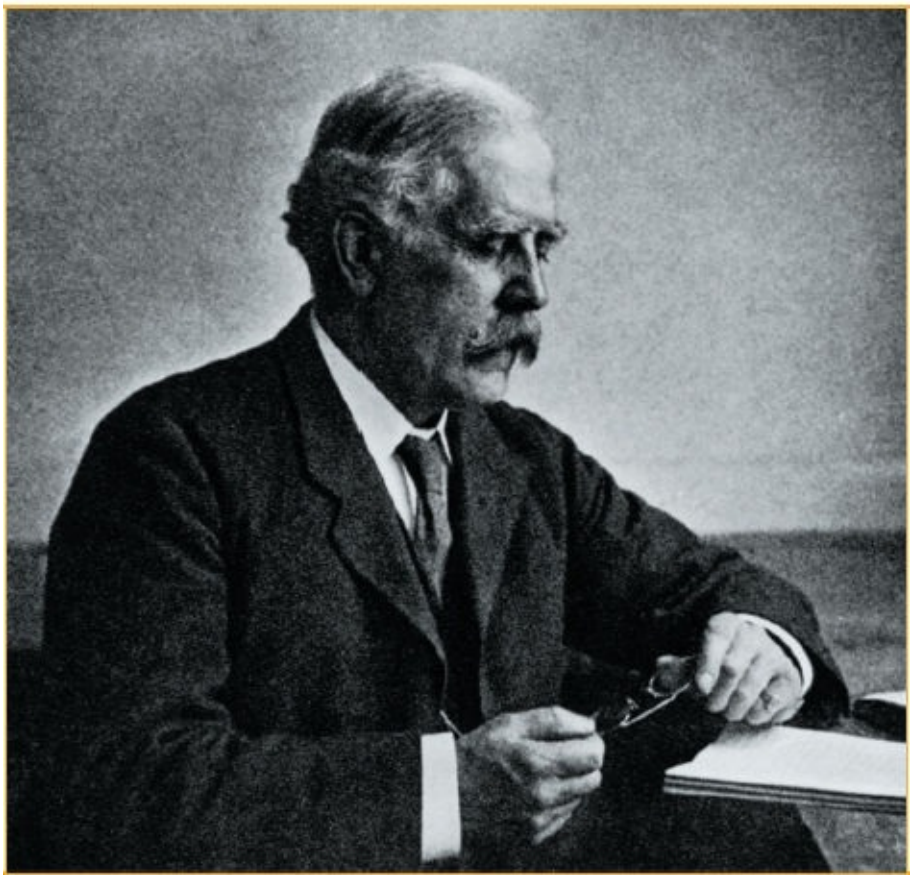


FIGURA 6.9 O médico britânico Archibald Garrod (1857-1936). (Science Photo Library/Science Source.)

Vias bioquímicas de síntese em Neurospora

O estudo de George Beadle e Edward Tatum foi um marco na década de 1940, pois não apenas esclareceu o papel dos genes, mas também demonstrou a interação dos genes nas vias bioquímicas. Eles posteriormente receberam um Prêmio Nobel por seu estudo, que marca o início de toda a biologia molecular. Beadle e Tatum realizaram seu trabalho com o fungo haploide *Neurospora*, que conhecemos nos capítulos anteriores. Seu plano era investigar o controle genético da química celular. No que se tornou a abordagem genética direta padrão, eles irradiaram pela primeira vez células de *Neurospora* para produzir mutações e em seguida testaram culturas crescidas a partir de ascosporos em relação aos fenótipos mutantes de interesse relevantes para a função bioquímica. Eles

observaram diversos mutantes que apresentavam nutrição defeituosa. Especificamente, esses mutantes eram mutantes auxotróficos, do tipo descrito em relação às bactérias no Capítulo 5. Enquanto o fungo *Neurospora* do tipo selvagem consegue utilizar a sua bioquímica celular para sintetizar virtualmente todos os seus componentes celulares a partir dos nutrientes inorgânicos e uma fonte de carbono no meio, os mutantes auxotróficos não conseguem. Para crescerem, os referidos mutantes necessitam que seja fornecido um nutriente (um nutriente que um fungo do tipo selvagem é capaz de sintetizar para si próprio), sugerindo que o mutante é defeituoso em relação a alguma etapa de síntese normal.

Como sua primeira etapa, Beadle e Tatum confirmaram que cada mutação que gerava uma exigência de nutriente era herdada como uma mutação monogênica, tendo em vista que cada uma delas fornecida uma proporção de 1:1 quando cruzada com um tipo selvagem. Deixando que *aux* represente uma mutação auxotrófica,

$$+ \times aux$$
$$\downarrow$$
$$\text{progênie: } \tfrac{1}{2} + \text{ e } \tfrac{1}{2}\, aux$$

Sua segunda etapa foi classificar a exigência nutricional específica de cada auxotrófico. Alguns cresceriam apenas se fosse fornecida prolina; outros, metionina; outros, piridoxina; outros, arginina; e assim por diante. Beadle e Tatum decidiram focar nos auxotróficos para arginina. Eles observaram que os genes que tinham sofrido mutação e originaram auxotróficos para arginina mapeavam em três *loci* diferentes em três cromossomos em separado. Iremos denominar os genes nos três *loci* de *arg-1*, *arg-2*, e *arg-3*. Uma conquista-chave foi a descoberta de Beadle e Tatum de que os auxotróficos em relação a cada um dos três *loci* diferiam em sua resposta aos compostos estruturalmente correlatos ornitina e citrulina (Figura 6.10). Os mutantes *arg-1* cresceram quando receberam suprimento de qualquer uma das substâncias químicas ornitina, citrulina ou arginina. Os mutantes *arg-2* cresceram quando receberam arginina ou citrulina,

mas não ornitina. Os mutantes *arg-3* cresceram apenas quando a arginina foi fornecida. Esses resultados estão resumidos na Tabela 6.1.

Já se sabia que as enzimas celulares interconvertem os referidos compostos correlatos. Com base nas propriedades dos mutantes *arg*, Beadle e Tatum e seus colegas propuseram uma via bioquímica em relação às referidas conversões em *Neurospora*:

$$\text{precursor} \xrightarrow{\text{enzima X}} \text{ornitina} \xrightarrow{\text{enzima Y}} \text{citrulina} \xrightarrow{\text{enzima Z}} \text{arginina}$$

Essa via explica bem as três classes de mutantes demonstradas na Tabela 6.1. Sob o modelo, os mutantes *arg-1* não apresentam a enzima X e, assim, são incapazes de converter o precursor em ornitina como a primeira etapa na produção de arginina. Entretanto, eles apresentam as enzimas normais Y e Z e, assim, os mutantes *arg-1* são capazes de produzir arginina se receberem suprimento de ornitina ou citrulina. De modo semelhante, os mutantes *arg-2* não apresentam a enzima Y, e os mutantes *arg-3* não apresentam a enzima Z. Portanto, presume-se que uma mutação em um gene em particular interfira na produção de uma única enzima. A falta da enzima cria um bloqueio em alguma via bioquímica de síntese. O bloqueio pode ser evitado por meio do fornecimento, para as células, de qualquer composto que normalmente vem após o bloqueio na via.

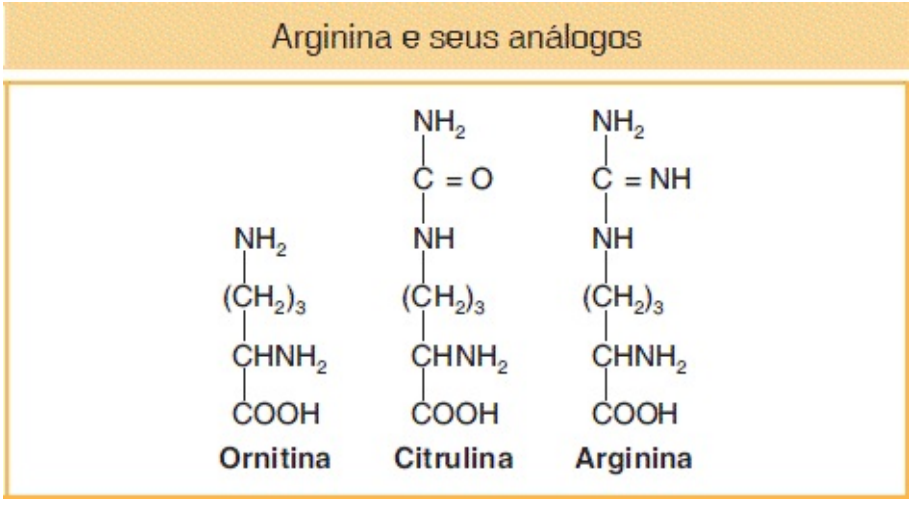


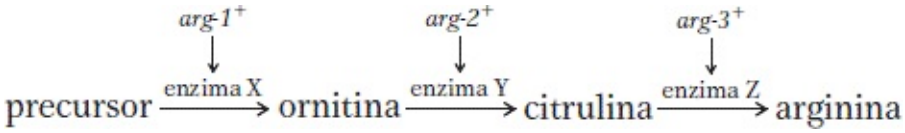
FIGURA 6.10 As estruturas químicas da arginina e dos compostos estruturalmente correlatos citrulina e ornitina.

Tabela 6.1 Crescimento de mutantes *arg* em resposta aos suplementos.

Mutante	Suplemento		
	Ornitina	**Citrulina**	**Arginina**
arg-1	+	+	+
arg-2	—	+	+
arg-3	—	—	+

Nota: um sinal de mais significa crescimento; um sinal de menos significa ausência de crescimento.

Agora, podemos diagramar um modelo bioquímico mais completo:



Esse modelo brilhante, que inicialmente era conhecido como a *hipótese um gene—uma enzima*, foi a fonte da primeira percepção excitante sobre as funções dos genes: os genes de algum modo eram responsáveis pela função das enzimas, e cada gene aparentemente controlava uma enzima específica em uma série de etapas interconectadas em uma via bioquímica. Outros pesquisadores obtiveram resultados semelhantes em relação a outras vias de síntese e a hipótese logo conquistou a aceitação geral. Também se observou que todas as proteínas, sejam enzimas ou não, são codificadas por genes e, assim, a frase foi refinada para se tornar a **hipótese um gene–um polipeptídio**. (Relembre que um polipeptídio é o tipo mais simples de proteína, uma cadeia única de aminoácidos.) Logo se tornou claro que um gene codifica a *estrutura física* de uma proteína, que, por sua vez,

dita a sua função. A hipótese de Beadle e Tatum se tornou um dos grandes conceitos de unificação em biologia, tendo em vista que proporcionou uma ponte que uniu as duas principais áreas de pesquisa, genética e bioquímica.

Devemos adicionar, entre parênteses, que embora a grande maioria dos genes codifique proteínas, atualmente sabe-se que algumas codificam RNA que apresentam funções especiais. Todos os genes são transcritos para produzir RNA. Os genes que codificam proteínas são transcritos em RNA mensageiro (mRNA), que em seguida é traduzido em proteína. Entretanto, o RNA codificado por uma minoria de genes nunca é traduzido em proteína, tendo em vista que o próprio RNA apresenta uma função única. Esses são denominados **RNA funcionais**. Alguns exemplos são os RNA transportador, RNA ribossômicos e pequenos RNA citoplasmáticos — saberemos mais a respeito deles nos capítulos posteriores.

> CONCEITO-CHAVE A síntese química nas células ocorre por meio de vias de etapas sequenciais catalisadas por enzimas. Os genes que codificam as enzimas de uma via específica constituem um subconjunto do genoma funcionalmente interativo.

Interação gênica em outros tipos de vias bioquímicas

A noção de que os genes interagem por meio de vias bioquímicas é uma noção poderosa, que apresenta aplicação em todos os organismos. A via da arginina de *Neurospora* é um exemplo de uma via de síntese, uma cadeia de conversões enzimáticas que sintetiza moléculas essenciais. Podemos estender novamente a ideia para um caso humano já introduzido, a doença fenilcetonúria (PKU), que é causada por um alelo autossômico recessivo. Essa doença resulta da incapacidade de converter a fenilalanina em tirosina. Como um resultado do bloqueio, a fenilalanina se acumula e é espontaneamente convertida em um composto tóxico, o ácido fenilpirúvico. O gene da PKU é parte de uma via metabólica como a via da arginina de *Neurospora*, e parte dela está demonstrada na Figura 6.11. A ilustração inclui diversas outras doenças causadas por

bloqueios em etapas dessa via (incluindo alcaptonúria, a doença investigada por Garrod).

Outro tipo de via bioquímica é a *via de transdução de sinal*. Esse tipo de via é uma cadeia de sinais complexos do ambiente para o genoma e de um gene para outro. Essas vias são cruciais para a função adequada de um organismo. Uma das vias de transdução de sinal mais bem-compreendidas surgiu a partir de uma análise genética da resposta reprodutiva de levedura. Dois tipos reprodutivos, determinados pelos alelos MATa e MATα, são necessários para que ocorra o cruzamento em levedura. Quando uma célula se encontra na presença de outra célula de tipo reprodutivo oposto, ela é submetida a uma série de alterações no formato e no comportamento como preparo para se reproduzir. A resposta reprodutiva é acionada por uma via de transdução de sinal que requer a ação sequencial de um conjunto de genes. Esse conjunto de genes foi descoberto por meio de uma análise de interação padrão de mutantes com resposta reprodutiva aberrante (a maior parte era estéril). As etapas foram unidas por meio da utilização das abordagens na próxima seção. O sinal que move esse conjunto de coisas é um feromônio (hormônio) reprodutivo liberado pelo tipo reprodutivo oposto; o feromônio se liga a um receptor de membrana, que é acoplado a uma proteína G dentro da membrana e que ativa a proteína. A proteína G, por sua vez, promove uma série de fosforilações sequenciais de proteínas, denominada cascata de quinases. Finalmente, a cascata ativa a transcrição de um conjunto de genes específicos de reprodução que possibilitam que a célula se reproduza. Uma mutação em qualquer uma dessas etapas pode comprometer o processo reprodutivo.

As *vias do desenvolvimento* compreendem as etapas por meio das quais um zigoto se torna um organismo adulto. Esse processo envolve muitas etapas geneticamente controladas, incluindo o estabelecimento dos eixos anteroposterior e dorsoventral, estabelecendo o plano corporal básico dos órgãos, e a diferenciação e a movimentação teciduais. Essas etapas podem requerer regulação gênica e transdução de sinal. As vias do desenvolvimento serão abordadas em detalhes no Capítulo 13, mas a interação de genes nessas vias é analisada do mesmo modo, conforme veremos em seguida.

Uma via bioquímica de síntese
e doenças correlacionadas

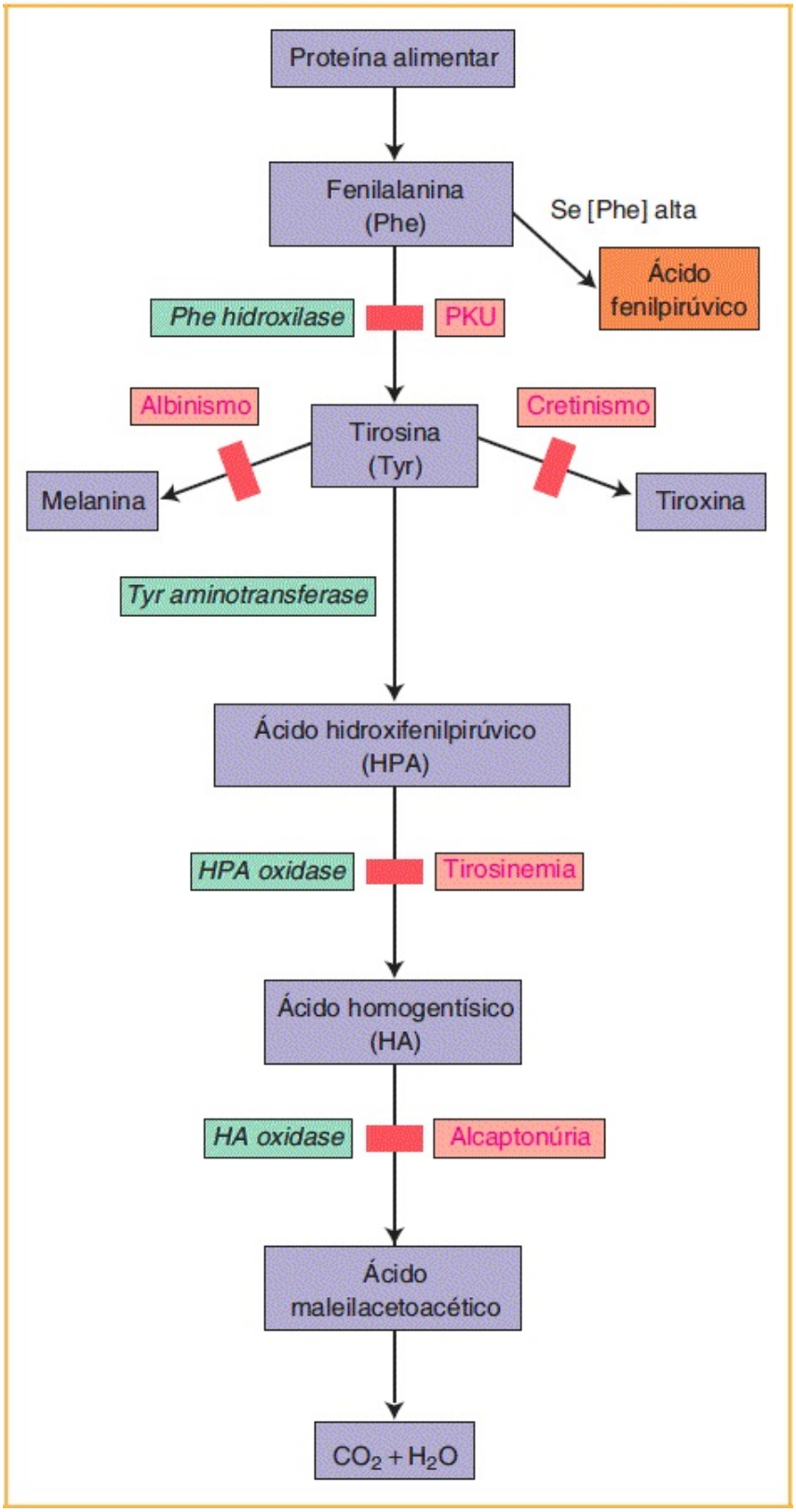


FIGURA 6.11 Uma seção da via metabólica da fenilalanina em seres humanos, incluindo doenças associadas a bloqueios enzimáticos. A fenilcetonúria é produzida quando a enzima fenilalanina hidroxilase apresenta mau funcionamento. O acúmulo de fenilalanina resulta em aumento de ácido fenilpirúvico, que interfere no desenvolvimento do sistema nervoso.

6.3 Inferência das interações gênicas

A abordagem genética que revela os genes interagindo em relação a uma propriedade biológica específica é, resumidamente, como segue:

Etapa 1. Obtenha muitos mutantes monogênicos e teste em relação à dominância.

Etapa 2. Teste os mutantes em relação ao alelismo — eles estão em um *locus* ou em diversos *loci*?

Etapa 3. Combine os mutantes em pares para formar **mutantes duplos**, para verificar se os genes interagem.

A interação gênica é inferida a partir do fenótipo do mutante duplo: se os genes interagem, então o fenótipo difere da simples combinação de ambos os fenótipos monogênicos mutantes. Se os alelos mutantes de diferentes genes interagem, então inferimos que os genes do tipo selvagem também normalmente interagem. Em casos nos quais os dois mutantes interagem, com frequência resultará uma proporção mendeliana modificada de 9:3:3:1.

Um procedimento que deve ser realizado antes de testar as interações é determinar se cada mutação é de um *locus* diferente (etapa 2 anterior). A triagem do mutante pode não intencionalmente ter favorecido determinados genes. Portanto, o conjunto de *loci* gênicos precisa ser definido, conforme demonstrado na próxima seção.

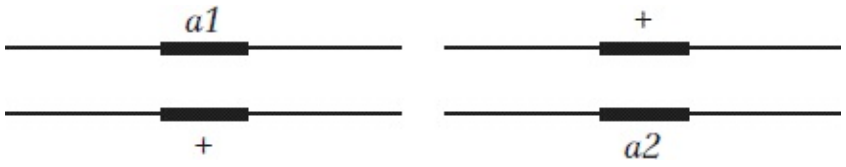
Segregação de mutantes com a utilização do teste de complementação

Como é possível decidir se duas mutações pertencem ao mesmo gene? Existem diversos modos. Primeiramente, cada alelo mutante pode ser mapeado. Em seguida, se duas mutações forem mapeadas em dois *loci* cromossômicos

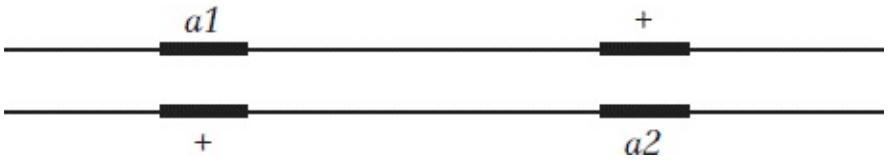
diferentes, provavelmente elas são de genes diferentes. Entretanto, essa abordagem é demorada em um grande conjunto de mutações. Uma abordagem mais rápida, utilizada com frequência, é o **teste de complementação**.

Em um diploide, o teste de complementação é realizado por meio do intercruzamento de dois indivíduos que são homozigotos em relação a diferentes mutações recessivas. A próxima etapa é observar se a progênie apresenta o fenótipo do tipo selvagem. Se a progênie for do tipo selvagem, as duas mutações recessivas estão obrigatoriamente em genes *diferentes*, tendo em vista que os respectivos alelos do tipo selvagem proporcionam função do tipo selvagem. Nesse caso, diz-se que as duas mutações se *complementaram*. Aqui, denominaremos os genes *a1* e *a2*, de acordo com seus alelos mutantes. Podemos representar os heterozigotos como segue, dependendo de os genes estarem no mesmo cromossomo ou estarem em cromossomos diferentes:

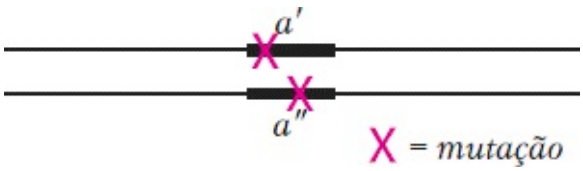
Cromossomos diferentes:



Mesmo cromossomo (demonstrado na configuração trans):



Entretanto, se a progênie *não* for do tipo selvagem, as mutações recessivas têm de ser alelos do mesmo gene. Tendo em vista que ambos os alelos do gene são mutantes, não existe um alelo do tipo selvagem para auxiliar na distinção entre dois alelos mutantes diferentes de um gene cujo alelo do tipo selvagem é a^+. Esses alelos podem apresentar sítios mutantes diferentes no mesmo gene, mas ambos seriam não funcionais. O heterozigoto *a'/a''* seria:



No nível operacional, a **complementação** é definida como a produção de um fenótipo do tipo selvagem quando dois genomas haploides que contêm mutações recessivas diferentes estão unidos na mesma célula.

Ilustraremos o teste de complementação com um exemplo de plantas campânula (gênero *Campanula*). A cor da flor do tipo selvagem dessa planta é azul. Presumiremos que, a partir de uma caça a uma mutante, tenhamos obtido três mutantes com pétalas brancas e que elas estejam disponíveis como linhagens homozigotas puras. Todas apresentam o mesmo aspecto e, assim, *a priori* não sabemos se elas são geneticamente idênticas. Denominaremos as linhagens mutantes $, £ e ¥ para evitar qualquer simbolismo com a utilização de letras, que pode implicar dominância. Quando cruzada com o tipo selvagem, cada mutante fornece os mesmos resultados na F_1 e na F_2, como segue:

$ branca × Azul → F_1, toda azul → F_2, $\frac{3}{4}$ azul, $\frac{1}{4}$ branca

£ branca × Azul → F_1, toda azul → F_2, $\frac{3}{4}$ azul, $\frac{1}{4}$ branca

¥ branca × Azul → F_1, toda azul → F_2, $\frac{3}{4}$ azul, $\frac{1}{4}$ branca

Flores da planta campânula (espécie de *Campanula*). (Gregory G. Dimijian/Science Source.)

Em cada caso, os resultados demonstram que a condição mutante é determinada pelo alelo recessivo de um único gene. Entretanto, existem três alelos de um gene, de dois genes ou de três genes? Tendo em vista que os mutantes são recessivos, a questão pode ser respondida por meio do teste de complementação, que indaga se os mutantes *complementam* uns aos outros.

Iremos entrecruzar os mutantes para testar a complementação. Presuma que os resultados do entrecruzamento dos mutantes $, £ e ¥ são como segue:

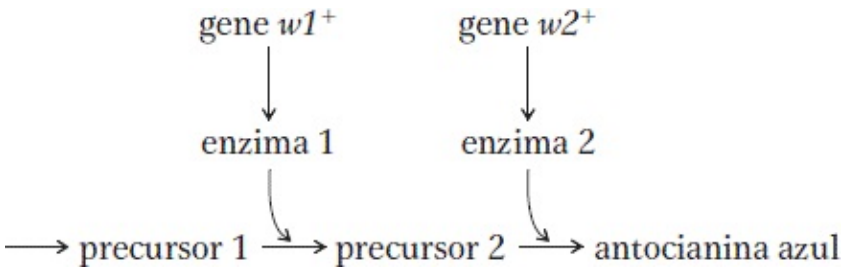
$ branca × £ branca → F_1, toda branca
$ branca × ¥ branca → F_1, toda azul
£ branca × ¥ branca → F_1, toda azul

A partir desse conjunto de resultados, podemos concluir que os mutantes $ e £ devem ser causados por alelos de um gene (digamos, *w1*), tendo em vista que eles não se complementam, mas ¥ deve ser causado por um alelo mutante de outro gene (*w2*), tendo em vista que ¥ complementa tanto $ quanto £.

> CONCEITO-CHAVE Quando dois alelos mutantes recessivos

> independentemente derivados, que produzem fenótipos recessivos semelhantes, não se complementam, eles têm de ser alelos do mesmo gene.

Como a complementação atua no nível molecular? A cor azul normal da flor da campânula é causada por um pigmento azul denominado *antocianina*. Os pigmentos são substâncias químicas que absorvem determinadas cores da luz; em relação à campânula, a antocianina absorve todos os comprimentos de onda com exceção do azul, que é refletido no olho do observador. Entretanto, essa antocianina é produzida a partir de precursores químicos que não são pigmentos; ou seja, eles não absorvem a luz de um comprimento de onda específico e simplesmente refletem de volta a luz branca do sol para o observador, proporcionando um aspecto branco. O pigmento azul é o produto final de uma série de conversões bioquímicas de não pigmentos. Cada etapa é catalisada por uma enzima específica codificada por um gene específico. Podemos explicar os resultados com uma via, como segue:



Uma mutação homozigota em qualquer dos genes levará ao acúmulo de um precursor que simplesmente tornará a planta branca. Agora as designações dos mutantes podem ser escritas como segue:

\$ $w1_{\$}/w1_{\$} \cdot w2^+/w2^+$
£ $w1_{£}/w1_{£} \cdot w2^+/w2^+$
¥ $w1^+/w1^+ \cdot w2_{¥}/w2_{¥}$

Entretanto, na prática, os símbolos subscritos seriam abandonados e os genótipos seriam escritos como segue:

$ *w1/w1 · w2⁺/w2⁺*

£ *w1/w1 · w2⁺/w2⁺*

¥ *w1⁺/w1⁺ · w2/w2*

Portanto, uma F_1 de $ × £ será:

w1/w1 · w2⁺/w2⁺

Essas plantas da F_1 apresentarão dois alelos defeituosos para *w1* e, portanto, serão bloqueadas na etapa 1. Embora a enzima 2 seja totalmente funcional, ela não apresenta um substrato sobre o qual atuar; assim, nenhum pigmento azul será produzido e o fenótipo será branco.

As plantas da F_1 de outros cruzamentos, entretanto, apresentarão os alelos do tipo selvagem para ambas as enzimas necessárias para levar os intermediários até o produto final azul. Seus genótipos serão:

w1⁺/w1 · w2⁺/w2

Portanto, observamos que a complementação é realmente um resultado da interação cooperativa de alelos do *tipo selvagem* dos dois genes. A Figura 6.12 resume a interação da complementação e da não complementação de mutantes brancos nos níveis genético e celular.

Em um organismo haploide, o teste de complementação não pode ser realizado por meio de entrecruzamento. Em fungos, um método alternativo une alelos mutantes para testar a complementação: a fusão que resulta em um **heterocário** (Figura 6.13). As células de fungo se fundem prontamente. Quando duas linhagens diferentes se fundem, os núcleos haploides das diferentes linhagens ocupam uma célula, que é o *heterocário* (grego; núcleos diferentes). Os núcleos em um heterocário em geral não se fundem. Em um sentido, essa condição é uma "mímica" de um diploide.

A base molecular da complementação genética

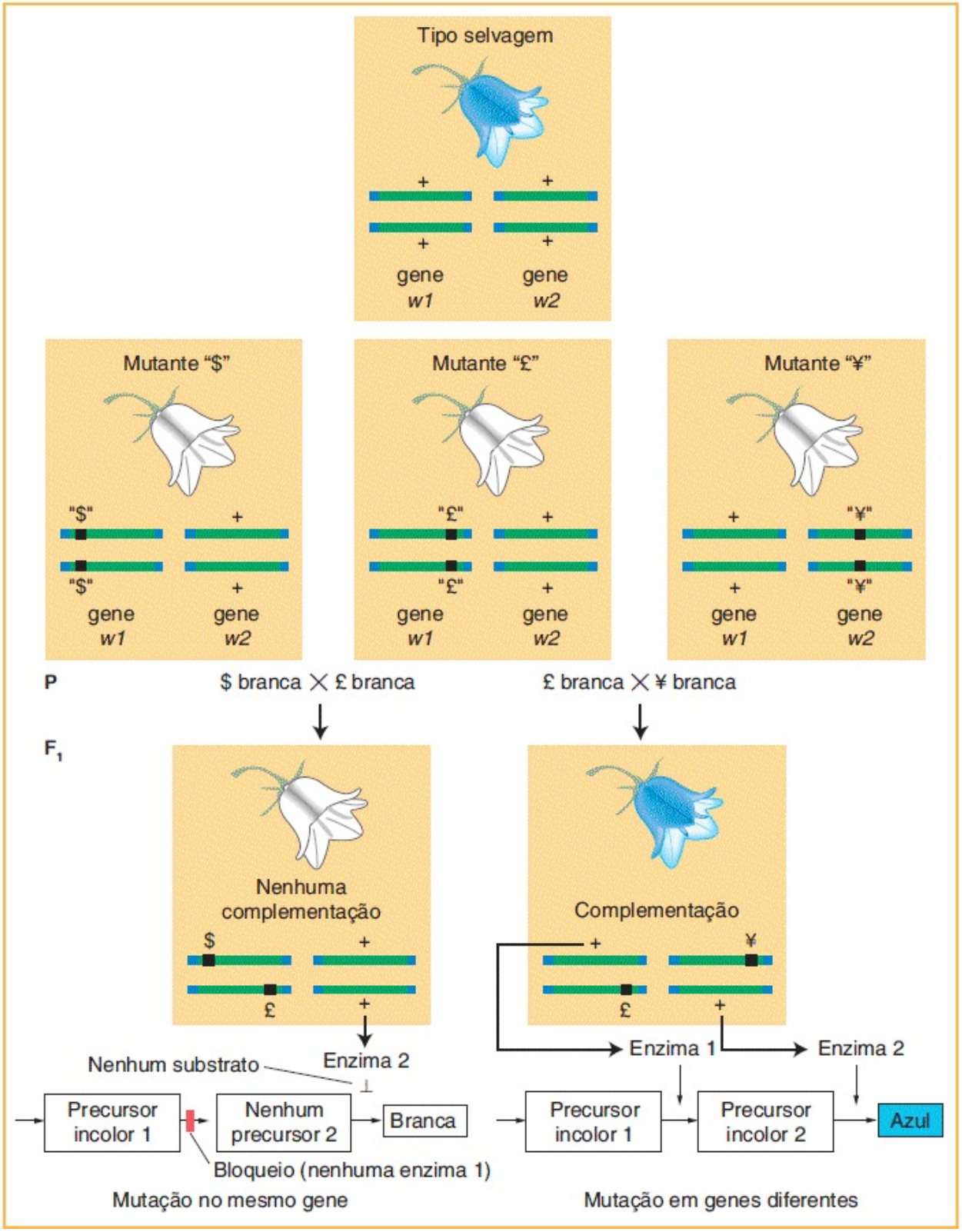


FIGURA 6.12 Três mutantes de campânula branca fenotipicamente idênticos ($, £ e ¥) são entrecruzados. As mutações no mesmo gene (tais como $ e £) não se complementam, tendo em vista que a F_1 apresenta um gene com dois alelos mutantes. A via é bloqueada e as flores são brancas. Quando as mutações estão em genes diferentes (tais como £ e ¥), existe complementação pelos alelos do tipo selvagem de cada gene no heterozigoto da F_1. O pigmento é sintetizado e as flores são azuis. (Qual você prevê que seja o resultado do cruzamento de $ e ¥?)

Presuma que, em linhagens diferentes, existam mutações em dois genes diferentes que conferem o mesmo fenótipo mutante — por exemplo, uma exigência de arginina. Denominaremos esses genes *arg-1* e *arg-2*. Os genótipos das duas linhagens podem ser representados como *arg-1* · *arg-2*$^+$ e *arg-1*$^+$ · *arg-2*. Essas duas linhagens podem ser fundidas para formar um heterocário com os dois núcleos em um citoplasma compartilhado:

O núcleo 1 é *arg-1* · *arg-2*$^+$
O núcleo 2 é *arg-1*$^+$ · *arg-2*

Tendo em vista que os produtos dos genes são criados em um citoplasma comum, os dois alelos do tipo selvagem podem exercer seu efeito dominante e colaborar para a produção de um heterocário de fenótipo do tipo selvagem. Em outras palavras, as duas mutações se complementam, justamente conforme ocorreria em um diploide. Se as mutações fossem de alelos do mesmo gene, não haveria complementação.

Análise de mutantes duplos de mutações aleatórias

Relembre que, para saber se dois genes interagem, precisamos avaliar o fenótipo do mutante duplo para verificar se ele é diferente da combinação de ambas as mutações únicas. O mutante duplo é obtido por meio de entrecruzamento. A F_1 é obtida como parte do teste de complementação; assim, com a presunção de que a complementação foi observada, sugerindo genes diferentes, a F_1 é autofecundada ou entrecruzada para se obter uma F_2 homozigota para ambas as mutações. Esse mutante duplo pode ser identificado em seguida ao se observarem as proporções mendelianas. Por exemplo, se uma proporção mendeliana 9:3:3:1 padrão for obtida, o fenótipo presente no único 1/16 da progênie representa o mutante duplo

(o “1” em 9:3:3:1). Nos casos de interação gênica, entretanto, o fenótipo do mutante duplo pode não ser distinto, mas corresponderá àquele dos mutantes únicos. Nesse caso, resultará uma proporção mendeliana modificada, tal como 9:3:4 ou 9:7.

A proporção mendeliana de 9:3:3:1 padrão é o caso mais simples, esperado se não houver interação gênica e se as duas mutações em teste estiverem em cromossomos diferentes. Essa proporção de 9:3:3:1 é a hipótese nula: qualquer proporção mendeliana modificada que represente um desvio dessa hipótese nula será informativa, como os exemplos a seguir irão demonstrar:

A proporção de 9:3:3:1 | Nenhuma interação gênica. Como um valor basal, iniciaremos com o caso no qual dois genes mutantes não interagem, uma situação na qual esperamos a proporção de 9:3:3:1. Vejamos a herança da cor da pele em cobras do milho. A cor natural da cobra é um padrão de repetição de camuflagem preta e laranja, conforme demonstrado na Figura 6.14 A. O fenótipo é produzido por dois pigmentos separados, ambos os quais estão sob controle genético. Um gene determina o pigmento laranja, e os alelos que consideraremos são o^+ (presença de pigmento laranja) e *o* (ausência de pigmento laranja). Outro gene determina o pigmento preto, e seus alelos são b^+ (presença de pigmento preto) e *b* (ausência de pigmento preto). Esses dois genes não estão ligados. O padrão natural é produzido pelo genótipo o^+/—; b^+/—. (O travessão representa a presença de qualquer alelo.) Uma cobra que é *o/o*; b^+/— é preta, tendo em vista que não apresenta o pigmento laranja (Figura 6.14 B), e uma cobra que é *o*+/—; *b/b* é laranja, tendo em vista que não apresenta o pigmento preto (Figura 6.14 C). O duplo homozigoto recessivo *o/o*; *b/b* é albino (Figura 6.14 D). Entretanto, observe que a cor rosa pálida do albino advém de um outro pigmento, a hemoglobina do sangue, que é visível através da pele dessa cobra quando os outros pigmentos estão ausentes. A cobra albina também demonstra claramente que existe outro elemento para o padrão de pigmentação da pele, além do pigmento: o motivo em repetição, no qual e ao redor do qual o pigmento é depositado.

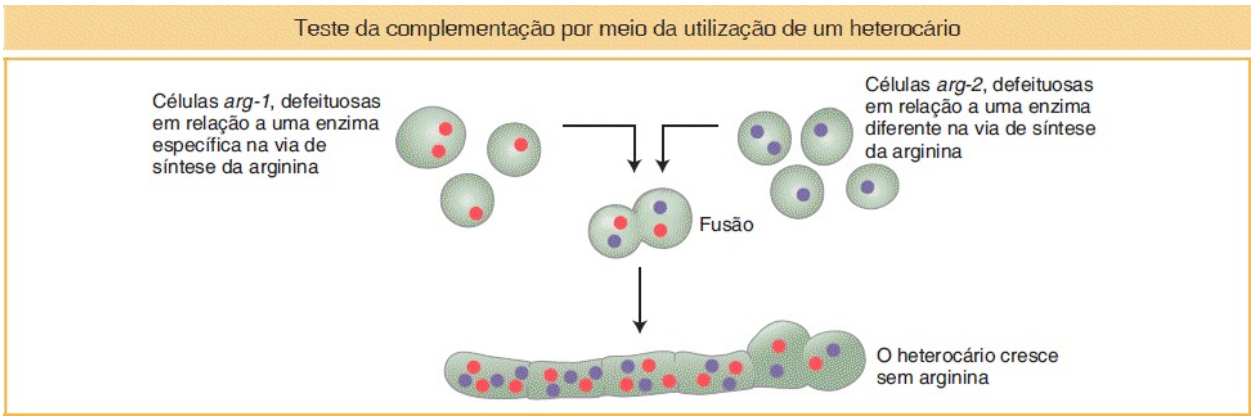


FIGURA 6.13 Um heterocário de *Neurospora* e fungos semelhantes mimetiza um estado diploide. Quando as células vegetativas se fundem, os núcleos haploides compartilham o mesmo citoplasma em um heterocário. Neste exemplo, os núcleos haploides com mutações em diferentes genes na via de síntese da arginina se complementam para produzir um *Neurospora* que deixa de necessitar de arginina.

FIGURA 6.14 Em cobras *Pantherophis guttatus*, combinações dos pigmentos laranja e preto determinam os quatro fenótipos demonstrados. **A.** Uma cobra camuflada preta e laranja do tipo selvagem sintetiza ambos os pigmentos preto e laranja. **B.** Uma cobra preta não sintetiza o pigmento laranja. **C.** Uma cobra laranja não sintetiza o pigmento preto. **D.** Uma cobra albina não sintetiza pigmento preto, nem laranja. (Anthony Griffiths.)

Se uma cobra laranja homozigota e uma cobra preta homozigota forem cruzadas, a F_1 é do tipo selvagem (camuflada), demonstrando a complementação:

$$\underset{\text{(laranja)}}{\female\ o^+/o^+\,;\,b/b} \times \underset{\text{(preta)}}{\male\ o/o\,;\,b^+/b^+}$$

$$\downarrow$$

$$F_1\ \ o^+/o\,;\,b^+/b$$
$$\text{(camuflada)}$$

Aqui, entretanto, uma F_2 demonstra uma proporção de 9:3:3:1 padrão:

♀ o^+/o ; b^+/b (camuflada)	×	♂ o^+/o ; b^+/b (camuflada)
↓		
F_2 9 $o^+/-$; $b^+/-$		(camuflada)
3 $o^+/-$; b/b		(laranja)
3 o/o ; $b^+/-$		(preta)
1 o/o ; b/b		(albina)

A proporção de 9:3:3:1 é produzida em virtude de os dois genes para pigmento atuarem de modo independente no nível celular.

precursor $\xrightarrow{b^+}$ pigmento preto
precursor $\xrightarrow{o^+}$ pigmento laranja
} camuflada

Se a presença de um mutante torna uma via bioquímica falha, a outra via ainda está ativa, produzindo a cor do outro pigmento. Ambas as vias falham apenas quando ambos os mutantes estão presentes, e não é produzido pigmento de nenhuma cor.

A proporção de 9:7 | Genes na mesma via bioquímica. A proporção da F_2 do cruzamento di-híbrido da campânula demonstra ambas as plantas azuis e brancas em uma proporção de 9:7. Como os referidos resultados podem ser explicados? A proporção de 9:7 é claramente uma modificação da proporção di-híbrida de 9:3:3:1, com 3:3:1 combinadas para perfazer 7; portanto, é inferido algum tipo de interação. O cruzamento das duas linhagens brancas e as gerações subsequentes podem ser representados como segue:

$w1/w1;w2^+/w2^+$ (branca) × $w1+/w1^+;w2/w2$ (branca)

↓

F_1 $w1^+/w1;w2^+/w2$ (azul)

$w1^+/w1;w2^+/w2$ × $w1^+/w1;w2^+/w2$

↓

F_2	9 $w1^+/-;w2^+/-$	(azul)	9
	3 $w1^+/-;w2/w2$	(branca)	7
	3 $w1/w1;w2^+/-$	(branca)	
	1 $w1/w1;w2/w2$	(branca)	

Claramente, nesse caso, o único modo por meio do qual é possível uma proporção de 9:7 é se o mutante duplo apresentar os mesmos fenótipos que os dois mutantes únicos. Portanto, a proporção modificada constitui um modo de identificar o fenótipo do mutante duplo. Além disso, os fenótipos idênticos dos mutantes únicos e duplos sugerem que cada alelo mutante controla uma etapa diferente na *mesma* via bioquímica. Os resultados demonstram que uma planta apresentará pétalas brancas se for homozigota em relação ao alelo mutante recessivo de *qualquer um* dos genes ou de *ambos* os genes. Para apresentar o fenótipo azul, uma planta deve apresentar no mínimo uma cópia do alelo dominante de ambos os genes, tendo em vista que ambos são necessários para completar as etapas sequenciais na via. Não importa qual esteja ausente, a mesma via bioquímica falha, produzindo o mesmo fenótipo. Portanto, três das classes genotípicas produzirão o mesmo fenótipo e, assim, em geral, resultam apenas dois fenótipos.

O exemplo nas campânulas envolveu etapas diferentes em uma via de síntese. Resultados semelhantes podem surgir a partir da regulação gênica. Um gene regulador com frequência atua por meio da produção de uma proteína que se liga a um sítio regulador *upstream* de um gene-alvo, facilitando a transcrição do gene (Figura 6.15). Na ausência da proteína reguladora, o gene-alvo será transcrito em níveis muito baixos, inadequados para as necessidades celulares. Cruzaremos uma linhagem pura *r/r* defeituosa para a proteína reguladora com uma linhagem pura *a/a* defeituosa para a proteína-alvo. O cruzamento é *r/r*; $a^+/a^+ \times r^+/r^+$; *a/a*. O

di-híbrido r^+/r; a^+/a demonstrará complementação entre os genótipos mutantes, tendo em vista que ambos, r^+ e a^+, estão presentes, possibilitando a transcrição normal do alelo do tipo selvagem. Quando autofecundados, os di-híbridos da F_1 também resultarão em uma proporção fenotípica de 9:7 na F_2:

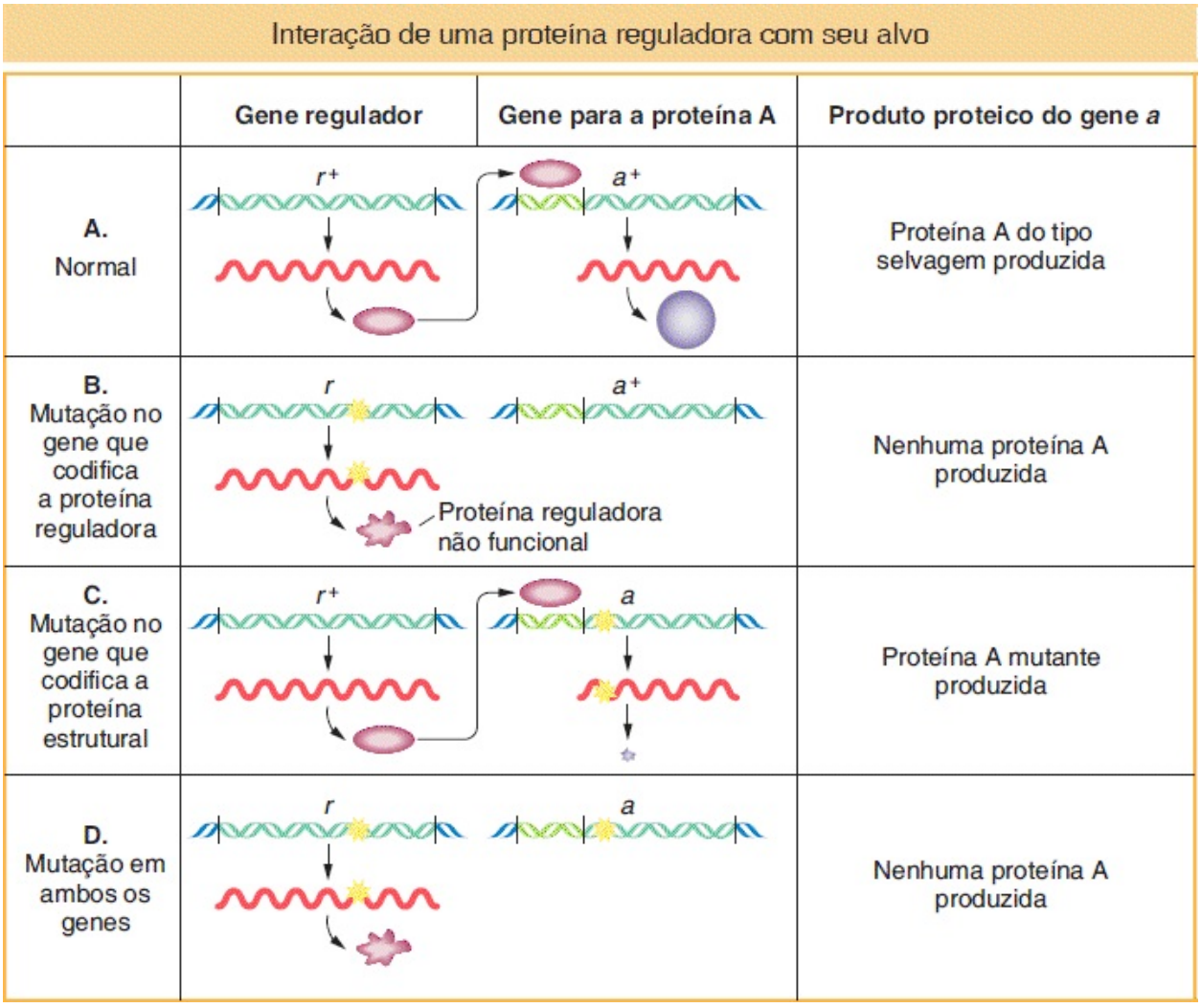


FIGURA 6.15 O gene r^+ codifica uma proteína reguladora, e o gene a^+ codifica uma proteína estrutural. Ambos têm de ser normais para que uma proteína estrutural funcional ("ativa") seja sintetizada.

Proporção	Genótipo	Proteína a^+ funcional	Proporção
$\frac{9}{16}$	$r^+/-\ ;\ a^+/-$	Sim	9
$\frac{3}{16}$	$r^+/-\ ;\ a/a$	Não	7
$\frac{3}{16}$	$r/r\ ;\ a^+/-$	Não	
$\frac{1}{16}$	$r/r\ ;\ a/a$	Não	

> CONCEITO-CHAVE Uma proporção de 9:7 na F_2 sugere interação de genes na mesma via; a ausência de qualquer função gênica leva à ausência do produto final da via.

A proporção de 9:3:4 | Epistasia recessiva. Uma proporção de 9:3:4 na F_2 sugere um tipo de interação gênica denominado **epistasia**. Essa palavra significa "predominância sobre", e se refere à situação na qual o mutante duplo demonstra o fenótipo de uma mutação, mas não de outra. A mutação que "predomina sobre" é *epistática*, enquanto a que "fica sob" é *hipostática*. A epistasia também resulta de os genes estarem na mesma via bioquímica. Em uma via de síntese simples, a mutação epistática é carreada por um gene que está mais *upstream* (mais inicialmente na via) do que o gene da mutação hipostática (Figura 6.16). O fenótipo mutante do gene *upstream* predomina, não importa o que esteja ocorrendo posteriormente na via.

Vejamos um exemplo a respeito da síntese do pigmento das pétalas na planta maria-olhos-azuis (*Collinsia parviflora*). A partir do tipo selvagem azul, iniciaremos com duas linhagens mutantes puras, uma com pétalas brancas (*w/w*) e a outra com pétalas magenta (*m/m*). Os genes *w* e *m* não estão ligados. A F_1 e a F_2 são como segue:

$w/w\,;\,m^+/m^+$ (branca) $\times$ $w^+/w^+\,;\,m/m$ (magenta)

F_1 $w^+/w\,;\,m^+/m$ (azul)

$\downarrow$

$w^+/w\,;\,m^+/m \times w^+/w\,;\,m^+/m$

$\downarrow$

F_2	9 $w^+/-\,;\,m^+/-$ (azul)	9
	3 $w^+/-\,;\,m/m$ (magenta)	3
	3 $w/w\,;\,m^+/-$ (branca)	4
	1 $w/w\,;\,m/m$ (branca)	

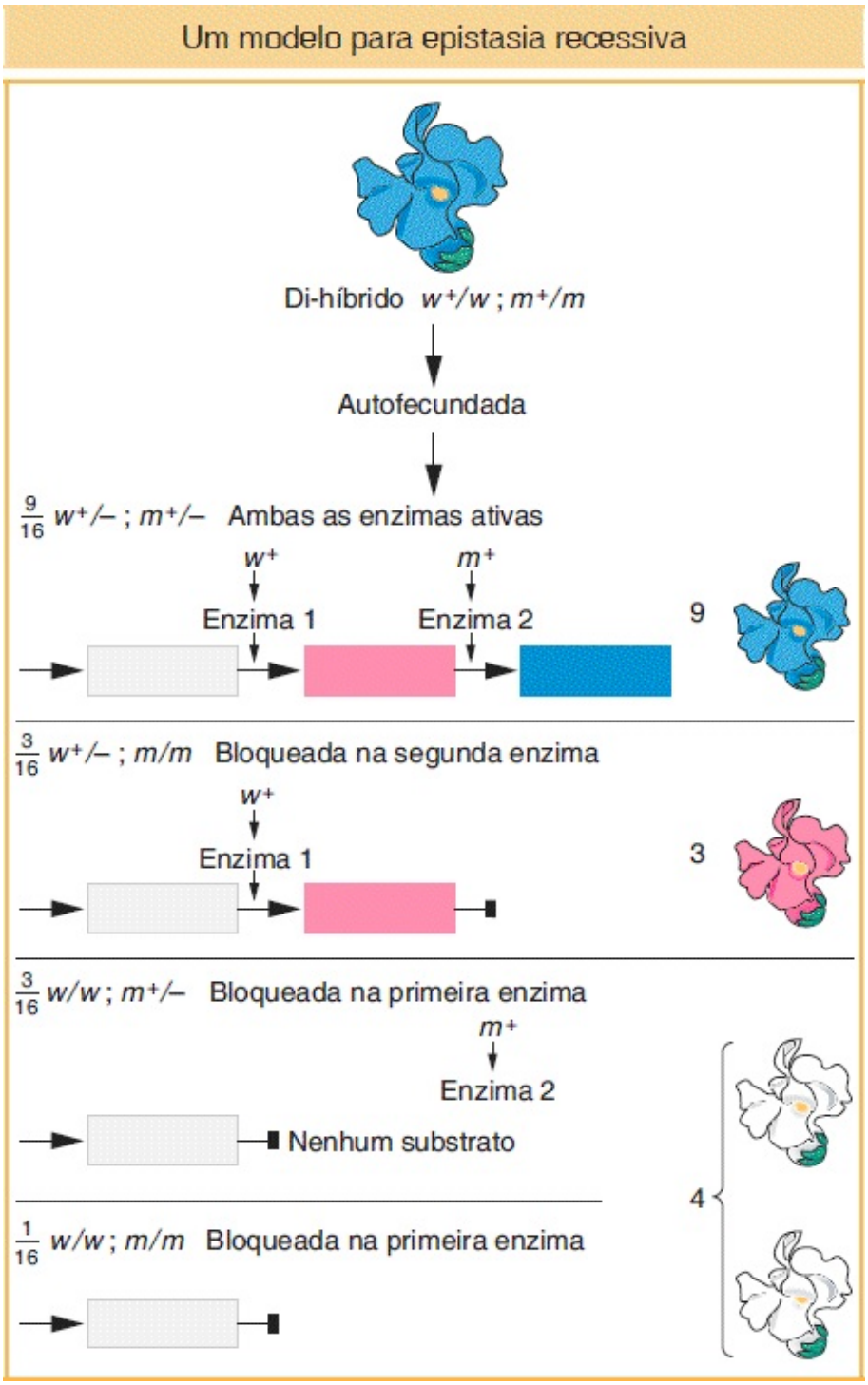


FIGURA 6.16 Alelos do tipo selvagem de dois genes (w^+ e m^+) codificam enzimas que catalisam etapas sucessivas na síntese de um pigmento azul das pétalas. As plantas homozigotas *m/m* produzem flores magenta, e as plantas homozigotas *w/w* produzem flores brancas. O mutante duplo *w/w*; *m/m* também produz flores brancas, indicando que o branco é epistático em relação ao magenta.

Na F_2, a proporção fenotípica de 9:3:4 é diagnóstica de epistasia recessiva. Assim como no caso precedente, novamente observamos que a proporção nos informa qual é obrigatoriamente o fenótipo do duplo mutante, tendo em vista que o componente $\frac{4}{16}$ da proporção deve ser um agrupamento de uma classe mutante única ($\frac{3}{16}$) mais a classe mutante dupla ($\frac{1}{16}$). Portanto, o mutante duplo expressa apenas um dos dois fenótipos mutantes; assim, por definição, branco tem de ser epistático em relação a magenta. (Para encontrar o mutante duplo dentro do grupo, as plantas brancas da F_2 deveriam ser individualmente submetidas ao cruzamento-teste.) Essa interação é denominada epistasia recessiva, tendo em vista que um fenótipo recessivo (branco) supera o outro fenótipo. A epistasia dominante será considerada na próxima seção.

No nível celular, podemos justificar a epistasia recessiva em *Collinsia* por meio do tipo de via a seguir (ver também Figura 6.16).

$$\text{incolor} \xrightarrow{\text{gene } w^+} \text{magenta} \xrightarrow{\text{gene } m^+} \text{azul}$$

Observe que a mutação epistática ocorre em uma etapa na via que leva ao pigmento azul; essa etapa é anterior à etapa que é bloqueada pela mutação mascarada.

Outro caso informativo de epistasia recessiva é a cor amarela da pelagem de alguns cães labrador retriever. Dois alelos, *B* e *b*, fazem referência às pelagens preta e marrom, respectivamente. Os dois alelos produzem melanina preta e marrom. O alelo *e* de outro gene é epistático em relação a esses alelos, proporcionando uma pelagem amarela (Figura 6.17). Portanto, ambos os genótipos *B/—*; *e/e* e *b/b*; *e/e* produzem um fenótipo amarelo, enquanto *B/—*; *E/—* e *b/b*; *E/—* são preto e marrom, respectivamente. Esse caso de epistasia *não* é causado por um bloqueio *upstream* na via que leva ao pigmento escuro. Cães amarelos podem produzir pigmento preto ou marrom, conforme pode ser observado em seus narizes e seus lábios. A ação do alelo *e* é evitar a deposição do pigmento nos pelos. Nesse caso, o gene epistático está *dowstream no desenvolvimento*; ele representa um tipo de alvo do desenvolvimento que tem de ser de genótipo *E* antes que o pigmento possa ser depositado.

Epistasia recessiva em virtude da mutação de pelagem amarela

FIGURA 6.17 Três cores de pelagem diferentes em labrador retriever. Dois alelos *B* e *b* de um gene de pigmento determinam (**A**) preto e (**B**) marrom, respectivamente. Em um gene em separado, *E* possibilita a deposição da cor na pelagem, e *e/e* impede a deposição, resultando (**C**) no fenótipo dourado. A parte **C** ilustra a epistasia recessiva. (Anthony Griffiths.)

> CONCEITO-CHAVE A epistasia é inferida quando um alelo mutante de um gene mascara a expressão de um alelo mutante de outro gene e em vez disso expressa o seu próprio fenótipo.

Em fungos, a análise de tétrades é útil na identificação de um mutante duplo. Por exemplo, um asco que contém metade dos seus produtos como do tipo selvagem deve conter mutantes duplos. Considere o cruzamento:

$$a \cdot b^{+} \times a^{+} \cdot b$$

Em alguma proporção da progênie, os alelos *a* e *b* segregarão juntos (um asco ditipo não parental). Uma referida tétrade demonstrará os fenótipos a seguir:

tipo selvagem $a^{+} \cdot b^{+}$ mutante duplo $a \cdot b$
tipo selvagem $a^{+} \cdot b^{+}$ mutante duplo $a \cdot b$

Portanto, o mutante duplo obrigatoriamente tem o genótipo não do tipo selvagem e pode ser avaliado de acordo. Se o fenótipo for o fenótipo *a*, então *b* está sendo sobrepujado; se o fenótipo for o fenótipo *b*, então *a* está sendo sobrepujado. Se ambos os fenótipos estiverem presentes, então não há epistasia.

A proporção de 12:3:1 | Epistasia dominante. Em dedaleiras (*Digitalis purpurea*), dois genes interagem na via que determina a coloração das pétalas. Os dois genes não estão ligados. Um gene afeta a intensidade do pigmento vermelho na pétala; o alelo *d* resulta na cor vermelho-clara observada em populações naturais de dedaleiras, enquanto *D* é um alelo mutante que produz a cor vermelho-escura (Figura 6.18). O outro gene determina em quais células o pigmento é sintetizado: o alelo *w* possibilita a síntese do pigmento por todas as pétalas assim como no tipo selvagem, mas o alelo mutante *W* confina a síntese do pigmento às pequenas manchas na garganta. Se autofecundarmos um di-híbrido *D/d*; *W/w*, a proporção da F_2 é como segue:

9 *D*/–; *W*/– (branca com manchas)	12
3 *d*/*d*; *W*/– (branca com manchas)	
3 *D*/–; *w*/*w* (vermelho-escuro)	3
1 *d*/*d*; *w*/*w* (vermelho-claro)	1

A proporção nos informa que o alelo dominante *W* é epistático, produzindo a proporção 12:3:1. O componente $\frac{12}{16}$ da proporção inclui obrigatoriamente a classe mutante dupla ($\frac{9}{16}$), que é claramente de fenótipo branco, estabelecendo a epistasia do alelo dominante *W*. Os dois genes atuam em uma via comum do desenvolvimento: *W* evita a síntese do pigmento vermelho, mas apenas em uma classe especial de células que constituem a área principal da pétala; a síntese é limitada às manchas da garganta. Quando a síntese é permitida, o pigmento pode ser produzido em concentrações altas ou baixas.

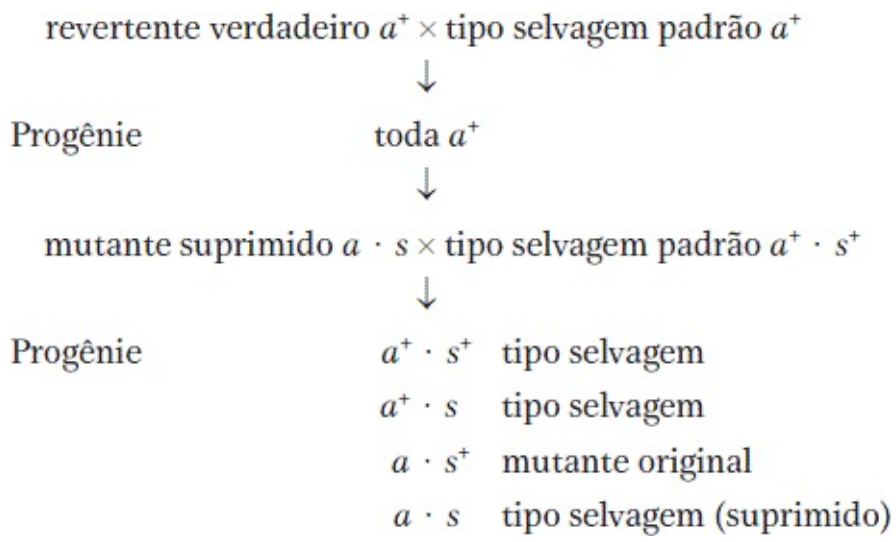
FIGURA 6.18 Em dedaleiras, *D* e *d* causam pigmentos escuros e claros, respectivamente, enquanto o *W* epistático restringe o pigmento às manchas na garganta. (Anthony Griffiths.)

Supressores. Não é fácil selecionar ou triar especificamente interações epistáticas e casos de epistasia devem ser construídos por meio da laboriosa combinação de duas mutações candidatas por vez. Entretanto, em relação ao nosso próximo tipo de interação gênica, o experimentador pode selecionar prontamente alelos mutantes de interesse. Um **supressor** é um alelo mutante de um gene que reverte o efeito de uma mutação de outro gene, resultando em um fenótipo do tipo selvagem ou próximo do tipo selvagem. A supressão implica que o gene-alvo e o gene supressor normalmente interagem em algum nível funcional em seus estados do tipo selvagem. Por exemplo, presuma que um alelo a^+ produza o fenótipo normal, enquanto um alelo mutante recessivo *a* resulte em uma anormalidade. Um alelo mutante recessivo *s* em outro gene suprime o efeito de *a* e, assim, o genótipo $a/a \cdot s/s$ apresentará o fenótipo do tipo selvagem (semelhante a a^+). Alelos supressores por vezes não apresentam efeito na ausência da outra mutação; em um referido caso, o fenótipo de $a^+/a^+ \cdot s/s$ seria do tipo selvagem.

Em outros casos, o alelo supressor produz o seu próprio fenótipo anormal.

A triagem de supressores é bem direta. Inicie com um mutante em algum processo de interesse, exponha esse mutante a agentes que causem mutação, tal como radiação de alta energia, e trie os descendentes em relação aos tipos selvagens. Em haploides tais como fungos, a triagem é realizada simplesmente por meio do plaqueamento de células mutagenizadas e da observação das colônias com fenótipos do tipo selvagem. Na maior parte, os tipos selvagens que são originados desse modo são meramente reversões do evento mutacional original e são denominados **revertentes**. Entretanto, alguns serão "pseudorrevertentes", mutantes duplos nos quais uma das mutações é um supressor.

Os estados revertente e suprimido podem ser distinguidos por meio de cruzamento apropriado. Por exemplo, em leveduras, os dois resultados seriam distinguidos como segue:



O surgimento do fenótipo mutante original identifica o genitor como um mutante suprimido.

Em diploides, os supressores produzem diversas proporções de F_2 modificadas, que são úteis na confirmação da supressão. Vejamos um exemplo da vida real de *Drosophila*. O alelo recessivo *pd* resulta na cor dos olhos roxa quando não suprimido. Um alelo *su* recessivo não apresenta por si próprio um

fenótipo detectável, mas suprime o alelo recessivo não ligado *pd*. Portanto, *pd/pd*; *su/su* tem o aspecto do tipo selvagem e apresenta olhos vermelhos. A análise a seguir ilustra o padrão de herança. Uma mosca com olhos roxos homozigota é cruzada com uma com olhos vermelhos homozigota que carreia o supressor.

$pd/pd;su^+/su^+$ (roxo) × $pd^+/pd^+;su/su$ (vermelho)

↓

F_1 todas $pd^+/pd;su^+/su$ (vermelho)

Autofecundada $pd^+/pd;su^+/su$ (vermelho) × $pd^+/pd;su^+/su$ (vermelho)

↓

F_2	9 $pd^+/-;su^+/-$	vermelho	13
	3 $pd^+/-;su/su$	vermelho	
	1 $pd/pd;su/su$	vermelho	
	3 $pd/pd;su^+/-$	roxo	3

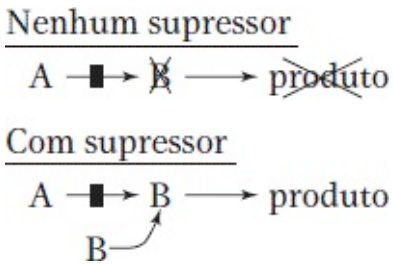
A proporção geral na F_2 é de 13 vermelhos:3 roxos. O componente de $\frac{13}{16}$ obrigatoriamente inclui o mutante duplo, que é claramente de fenótipo do tipo selvagem. Essa proporção é esperada a partir de um supressor recessivo que por si próprio não apresenta fenótipo detectável.

Por vezes a supressão é confundida com epistasia. Entretanto, a diferença-chave é que um supressor cancela a expressão de um alelo mutante e restaura o fenótipo do tipo selvagem correspondente. Além disso, com frequência apenas dois fenótipos segregam (assim como nos exemplos precedentes), em vez de três, como na epistasia.

Como os supressores atuam no nível molecular? Existem muitos mecanismos possíveis. Um tipo de supressão particularmente útil tem por base a ligação física de produtos gênicos na célula — por exemplo, ligação entre proteínas. Presuma que duas proteínas normalmente se combinem para proporcionar algum tipo de função celular. Quando uma mutação causa uma alteração no formato de uma proteína, ela deixa de se combinar à outra; portanto, a função é perdida (Figura

6.19). Entretanto, uma mutação supressora que causa uma alteração no formato compensatória na segunda proteína pode restaurar a combinação, portanto, a função normal. Nessa figura, se os genótipos fossem diploides que representam uma F_2 de um di-híbrido, então resultaria uma proporção de 14:2, tendo em vista que os únicos genótipos mutantes seriam $m/m \cdot s^+/s^+$ (1/16) e $m^+/m^+ \cdot s/s$ (1/16), totalizando 2/16. Se esse fosse um cruzamento di-híbrido haploide (tal como m^+ s^+ × m s), resultaria uma proporção de 1:1. A partir das proporções de supressores em geral, com frequência as proteínas que interagem podem ser deduzidas.

Alternativamente, quando uma mutação causa um bloqueio em uma via metabólica, o supressor encontra algum modo de superar o bloqueio — por exemplo, por meio do redirecionamento de intermediários similares para além do bloqueio da via. No exemplo a seguir, o supressor fornece um intermediário B para contornar o bloqueio.



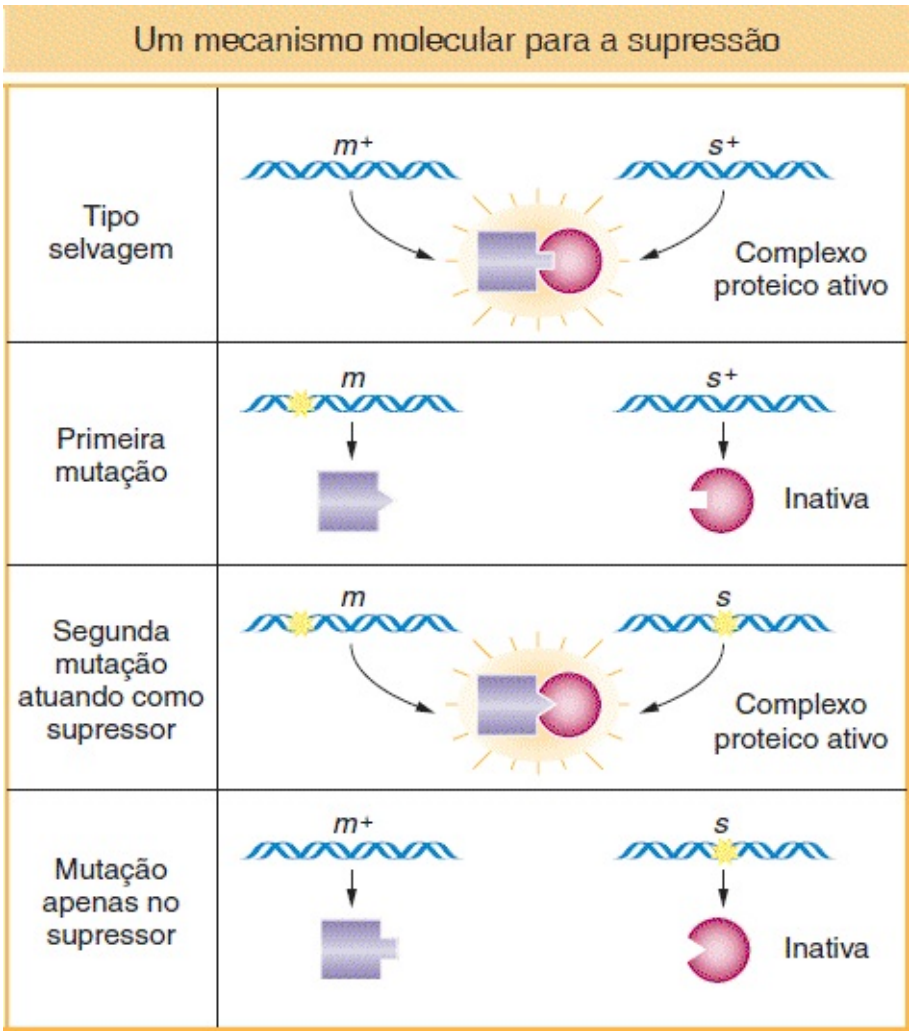


FIGURA 6.19 Uma primeira mutação altera o sítio de ligação de uma proteína, de modo que ela não consegue mais se ligar a um parceiro. Uma mutação supressora no parceiro altera o sítio de ligação, de modo que ambas as proteínas são capazes de se ligar novamente.

Em diversos organismos, foram encontrados *supressores sem sentido* — mutações nos genes de tRNA que resultam em um anticódon que se ligará a um códon de parada prematuro em uma sequência codificadora mutante. Portanto, o supressor possibilita que a tradução prossiga além do bloqueio anterior e produza uma proteína completa, em vez de uma proteína truncada. Essas mutações supressoras com frequência apresentam poucos efeitos sobre o fenótipo além da supressão.

> CONCEITO-CHAVE Alelos mutantes denominados supressores cancelam o efeito de um alelo mutante de outro gene, resultando em fenótipo do tipo selvagem.

Modificadores. Como a denominação sugere, uma mutação **modificadora** em um segundo *locus* altera o grau de expressão de um gene mutado no primeiro *locus*. Os genes reguladores fornecem uma ilustração simples. Assim como em um exemplo anterior, as proteínas reguladoras se ligam à sequência do DNA *upstream* do sítio de início da transcrição. Essas proteínas regulam o nível de transcrição. Na discussão sobre a complementação, consideramos uma mutação nula de um gene regulador que evitou quase completamente a transcrição. Entretanto, algumas mutações reguladoras alteram o nível de transcrição do gene-alvo, de modo que mais ou menos proteína é produzida. Em outras palavras, uma mutação em uma proteína reguladora pode diminuir ou aumentar o gene transcrito. Vejamos um exemplo que utiliza uma mutação *b* em um regulador que regula para menos, que afeta o gene *A* em um fungo, tal como levedura. Observamos o efeito de *b* sobre uma mutação *leaky* no gene *A*. Uma mutação *leaky* é uma mutação com algum nível baixo de função gênica. Cruzamos uma mutação gotejante *a* com a mutação reguladora *b*:

Mutante *leaky* $a \cdot b^+$ × Regulador ineficiente $a^+ \cdot b$

Progênie	Fenótipo
$a^+ \cdot b^+$	tipo selvagem
$a^+ \cdot b$	defeituoso (baixa transcrição)
$a \cdot b^+$	defeituoso (proteína A defeituosa)
	extremamente defeituoso (baixa transcrição da proteína

$a \cdot b$	defeituosa)

Portanto, a ação do modificador é observada no aparecimento de duas gradações de fenótipos mutantes *dentro* da progênie *a*.

Letais sintéticos. Em alguns casos, quando dois mutantes únicos viáveis são intercruzados, os mutantes duplos resultantes são letais. Em uma F_2 diploide, esse resultado seria manifestado como uma proporção de 9:3:3, tendo em vista que o mutante duplo (que seria o componente "1" da proporção) estaria ausente. Esses **letais sintéticos** podem ser considerados uma categoria especial de interação gênica. Eles podem apontar para tipos específicos de interações de produtos gênicos. Por exemplo, a análise do genoma revelou que a evolução produziu muitos sistemas duplicados dentro da célula. Uma vantagem desses sistemas duplicados pode ser o fornecimento de "cópias de segurança". Se houver mutações nulas nos genes em ambos os sistemas duplicados, então um sistema defeituoso não apresentará uma cópia de segurança, e o indivíduo não apresentará a função essencial e morrerá. Em outro caso, uma mutação *leaky* em uma etapa de uma via bioquímica pode causar a diminuição da velocidade da via, mas deixar função suficiente para a vida. Entretanto, se mutantes duplos são combinados, cada um com uma mutação *leaky* em uma etapa diferente, a via inteira é interrompida. Uma versão da última interação é de duas mutações em proteínas que interagem, conforme demonstrado na Figura 6.20.

Nas discussões anteriores sobre proporções mendelianas modificadas, todos os cruzamentos eram autofecundações de di-híbridos. Como um exercício, você pode querer calcular as proporções que seriam produzidas nos mesmos sistemas se fossem realizados cruzamentos-teste em vez de autofecundações.

> CONCEITO-CHAVE Uma diversidade de proporções da F_1 de 9:3:3:1 modificadas pode revelar tipos específicos de interação gênica.

Um resumo de algumas das proporções que revelam a interação gênica está demonstrado na Tabela 6.2.

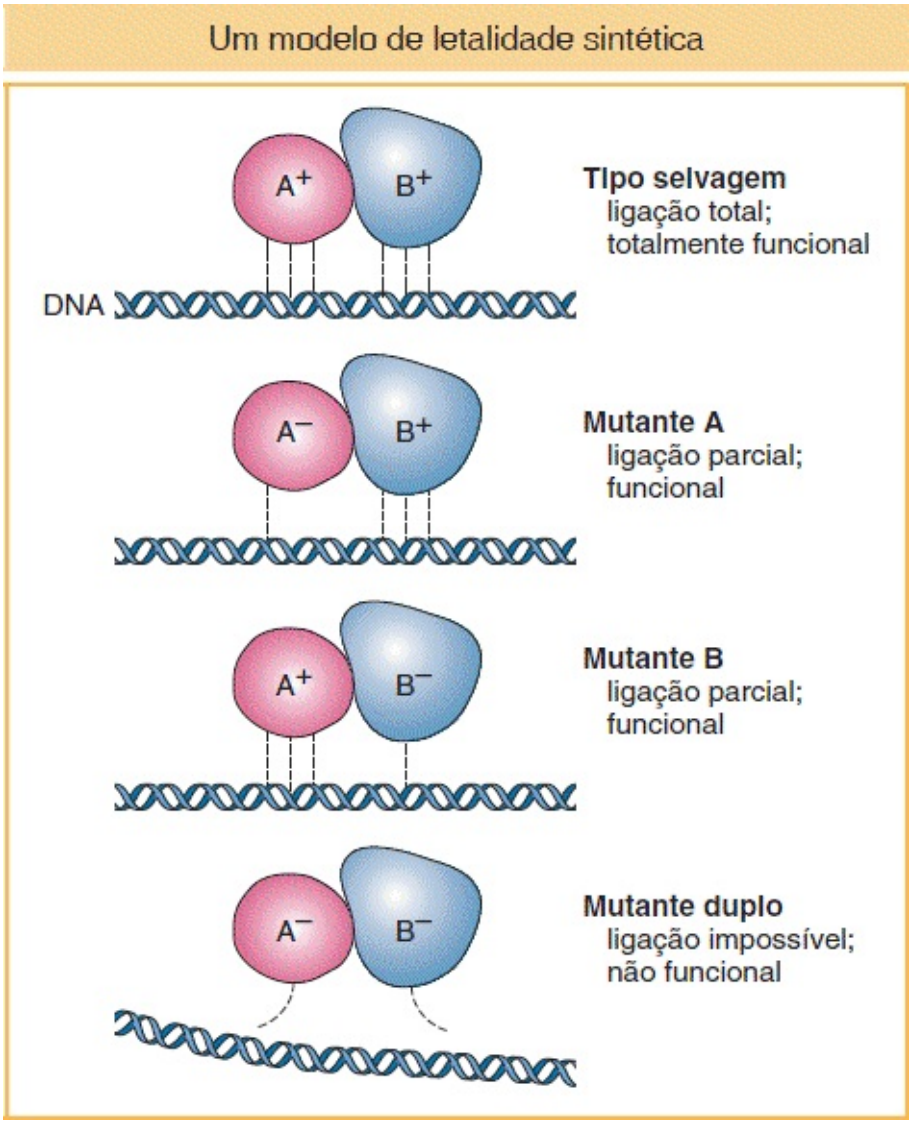
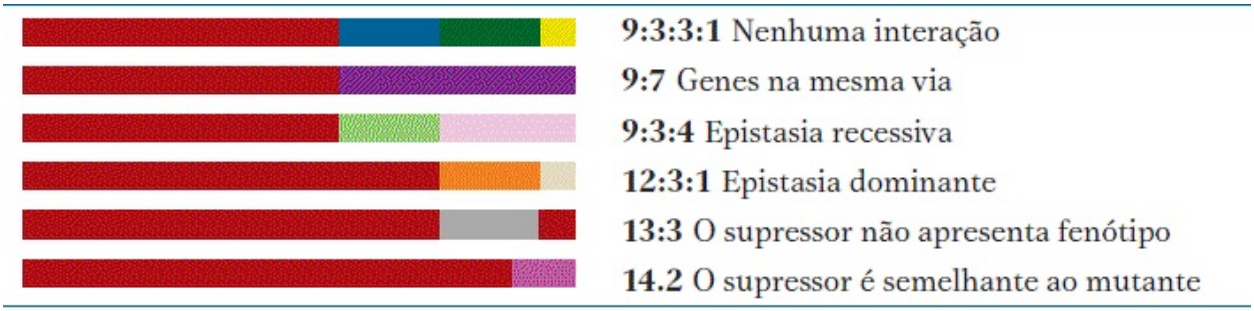


FIGURA 6.20 Duas proteínas em interação realizam alguma função essencial em algum substrato, tal como DNA, mas primeiramente precisam se ligar a ele. A ligação reduzida de qualquer proteína possibilita a permanência de algumas funções, mas a ligação reduzida de ambas é letal.

Tabela 6.2 Algumas proporções de F_2 modificadas.



Nota: algumas dessas proporções podem ser produzidas com outros mecanismos de interação.

6.4 Penetrância e expressividade

Na análise da herança monogênica, existe uma tendência natural de escolher mutantes que produzem proporções mendelianas claras. Nos referidos casos, podemos utilizar o fenótipo para distinguir os genótipos mutantes e do tipo selvagem com quase 100% de certeza. Nesses casos, dizemos que a mutação é 100% *penetrante* no fenótipo. Entretanto, muitas mutações demonstram penetrância *incompleta*: ou seja, nem todos os indivíduos com o genótipo expressam o fenótipo correspondente. Portanto, a **penetrância** é definida como a porcentagem de *indivíduos* com um determinado alelo que exibem o fenótipo associado àquele alelo.

Por que um organismo apresentaria um determinado genótipo e ainda assim não expressaria o fenótipo correspondente? Existem diversos motivos possíveis:

1. *A influência do ambiente.* Indivíduos com o mesmo genótipo podem demonstrar uma diversidade de fenótipos, dependendo do ambiente. A variação dos fenótipos em relação aos indivíduos mutantes e do tipo selvagem pode se sobrepor: o fenótipo de um indivíduo mutante que cresceu em um conjunto de circunstâncias pode corresponder ao fenótipo de um indivíduo do tipo selvagem que cresceu em um conjunto de circunstâncias diferente. Se essa correspondência ocorrer, o mutante não pode ser distinguido do tipo selvagem.
2. *A influência de interação com outros genes.* Modificadores não caracterizados, genes epistáticos ou supressores no restante do genoma

atuam para prevenir a expressão do fenótipo típico.

3. *A sutileza do fenótipo mutante*. Os efeitos sutis ocasionados pela ausência de uma função gênica podem ser de difícil mensuração em uma situação de laboratório.

Um exemplo típico com penetrância incompleta está demonstrado na Figura 6.21. Nesse heredograma humano, observamos um fenótipo herdado normalmente de modo dominante desaparecendo na segunda geração, apenas para reaparecer na próxima.

Outra medida para descrever a variação da expressão fenotípica é denominada **expressividade**. A expressividade mede o grau no qual um determinado alelo é expresso no nível fenotípico; ou seja, a expressividade mede a intensidade do fenótipo. Por exemplo, animais "marrons" (genótipo *b/b*) de diferentes estoques podem demonstrar intensidades muito diferentes do pigmento marrom, do claro até o escuro. Em relação à penetrância, a expressividade variável pode ocorrer em virtude da variação na constituição alélica do restante do genoma ou de fatores ambientais. A Figura 6.22 ilustra a distinção entre penetrância e expressividade. Um exemplo de expressividade variável em cães é observado na Figura 6.23.

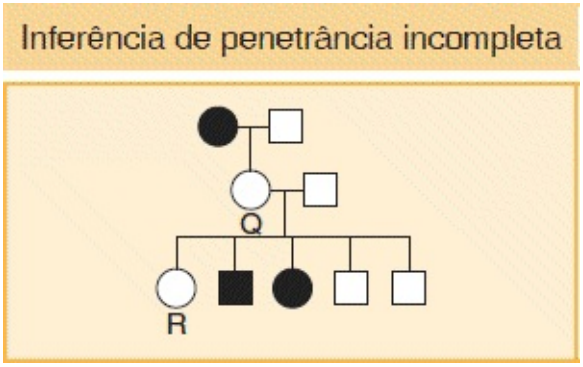


FIGURA 6.21 Neste heredograma humano de um alelo dominante que não é totalmente penetrante, a pessoa Q não demonstra o fenótipo, mas transmitiu o alelo dominante para no mínimo dois descendentes. Tendo em vista que alelo não é totalmente penetrante, o restante da progênie (p. ex., R) pode ter herdado o alelo dominante ou não.

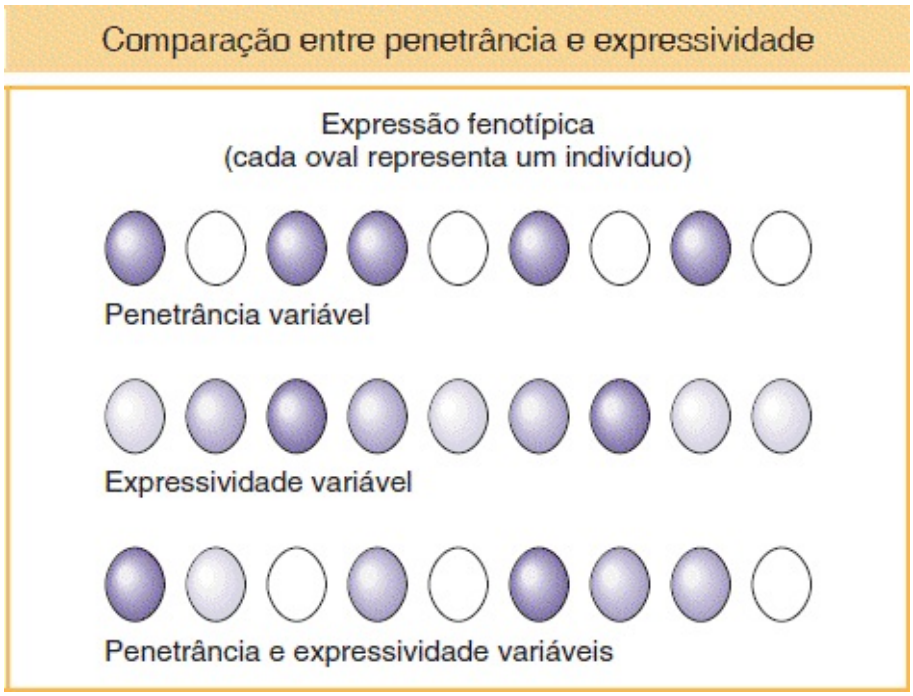


FIGURA 6.22 Presuma que todos os indivíduos demonstrados apresentam o mesmo alelo de um pigmento (*P*) e possuem o mesmo potencial de produzir o pigmento. Os efeitos do restante do genoma e do ambiente podem suprimir ou modificar a produção do pigmento em qualquer indivíduo. A cor indica o nível de expressão.

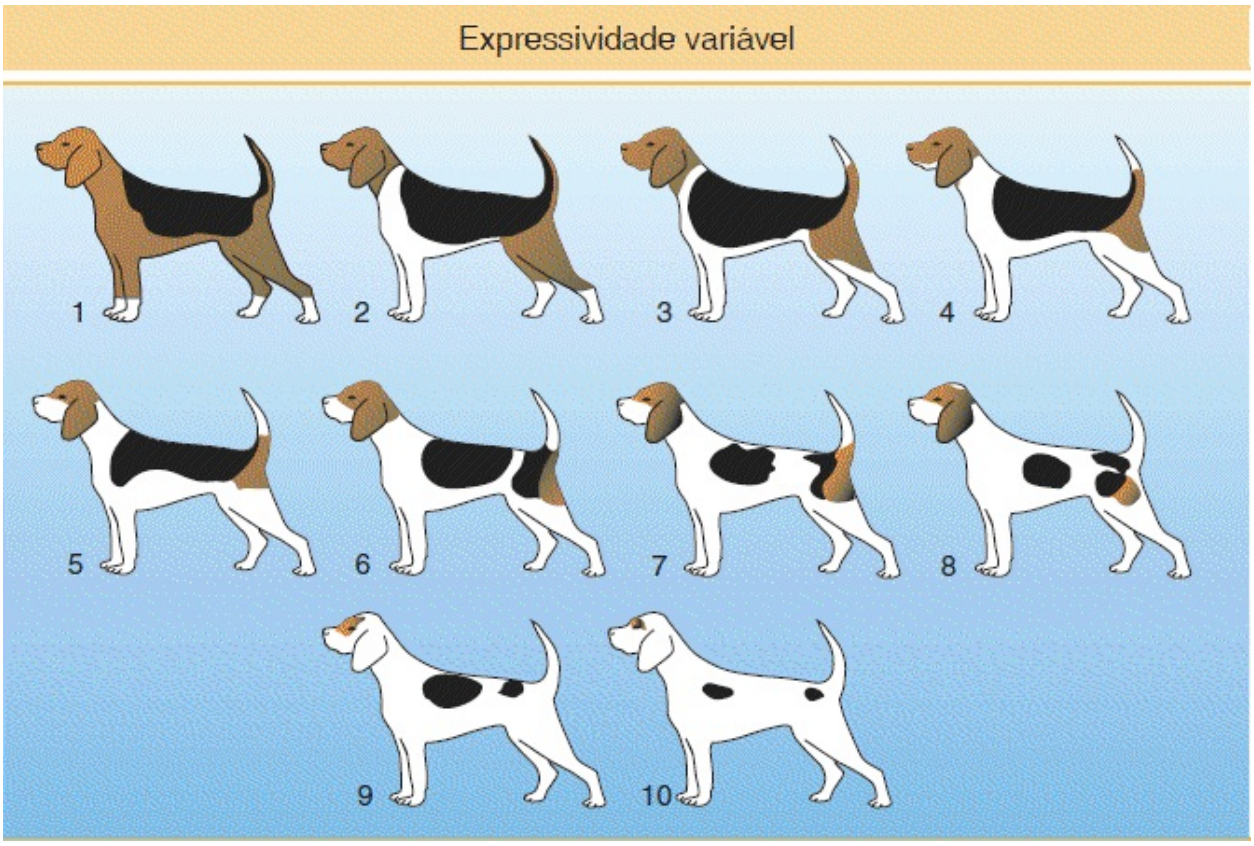


FIGURA 6.23 Dez gradações de padrões de manchas não pigmentadas em beagles. Cada um desses cães apresenta o alelo S^P, responsável pelas manchas não pigmentadas nos cães. A variação é causada pela variação em outros *loci*.

Os fenômenos de penetrância incompleta e expressividade variável podem tornar qualquer tipo de análise genética substancialmente mais difícil, incluindo a análise de heredogramas humanos e as previsões no aconselhamento genético. Por exemplo, esse com frequência é o caso em que um alelo que causa uma doença não é totalmente penetrante. Portanto, uma pessoa poderia apresentar o alelo, mas não demonstrar sinais da doença. Se esse for o caso, é difícil fornecer um claro laudo genético de saúde para uma pessoa em um heredograma de doença (p. ex., a pessoa R na Figura 6.21). Por outro lado, a análise de heredogramas por vezes consegue identificar pessoas que não expressam, mas que quase certamente apresentam um genótipo de doença (p. ex., o indivíduo Q na Figura 6.21). De modo semelhante, a expressividade variável pode complicar o aconselhamento, tendo em vista que pessoas com baixa expressividade poderiam ser erroneamente

diagnosticadas.

Embora a penetrância e a expressividade possam ser quantificadas, não obstante, elas representam situações “nebulosas”, tendo em vista que raramente é possível identificar fatores específicos que causam a variação sem pesquisas substanciais adicionais.

> CONCEITO-CHAVE Os termos *penetrância* e *expressividade* quantificam a modificação do efeito de um gene por meio da variação do ambiente e do *background* genético; eles medem, respectivamente, a porcentagem de casos nos quais o fenótipo é observado e a sua gravidade.

RESUMO

Um gene não atua isoladamente; em vez disso, ele atua em conjunto com muitos outros genes no genoma. Na análise genética direta, a dedução dessas interações complexas é um estádio importante da pesquisa. As mutações individuais são testadas primeiramente a respeito de suas relações de dominância, um tipo de interação alélica. As mutações recessivas com frequência resultam da haplossuficiência do alelo do tipo selvagem, enquanto as mutações dominantes com frequência são o resultado da haploinsuficiência do tipo selvagem ou da ação do mutante como um dominante negativo (um polipeptídio trapaceiro). Algumas mutações causam efeitos graves ou até mesmo morte (mutações letais). A letalidade de uma mutação homozigota recessiva é um modo de avaliar se um gene é essencial no genoma.

A interação de diferentes genes resulta de sua participação na mesma via bioquímica ou em vias de conexão de diversos tipos — síntese, transdução de sinal ou desenvolvimento. A dissecção genética das interações gênicas tem início por meio da reunião, por parte do experimentador, de mutantes que afetam um caractere de interesse. O teste de complementação determina se duas mutações recessivas distintas são de um gene ou de dois genes diferentes. Os genótipos mutantes são reunidos em um indivíduo da F_1, e se o fenótipo for mutante, então

não ocorreu nenhuma complementação e os dois alelos são obrigatoriamente do mesmo gene. Se o fenótipo for do tipo selvagem, então ocorreu complementação, e os alelos têm de ser de genes diferentes.

A interação dos diferentes genes pode ser detectada por meio do teste de mutantes duplos, tendo em vista que a interação alélica implica a interação de produtos gênicos no nível funcional. Alguns tipos-chave de interação são a epistasia, a supressão e a letalidade sintética. A epistasia é a substituição de um fenótipo mutante produzido por uma mutação por um fenótipo mutante produzido por mutação em outro gene. A observação de epistasia sugere uma via em comum, seja química ou do desenvolvimento. Um supressor é uma mutação de um gene que consegue restaurar o fenótipo do tipo selvagem em uma mutação em outro gene. Os supressores com frequência revelam proteínas ou ácidos nucleicos que interagem fisicamente. Algumas combinações de mutantes viáveis são letais, um resultado conhecido como letalidade sintética. Os *letais sintéticos* podem revelar uma diversidade de interações, dependendo da natureza das mutações.

Os diferentes tipos de interações gênicas produzem proporções di-híbridas de F_2 que são modificações do padrão de 9:3:3:1. Por exemplo, a epistasia recessiva resulta em uma proporção de 9:3:4.

Em termos mais gerais, a interação gênica e a interação do gene com o ambiente são reveladas por meio da penetrância variável (a capacidade de um genótipo de expressar a si próprio no fenótipo) e da expressividade (o grau quantitativo da manifestação fenotípica de um genótipo).

TERMOS-CHAVE

alelo letal
alelos múltiplos
alelo pleiotrópico
codominância
complementação
dominância incompleta
dominância total (completa)

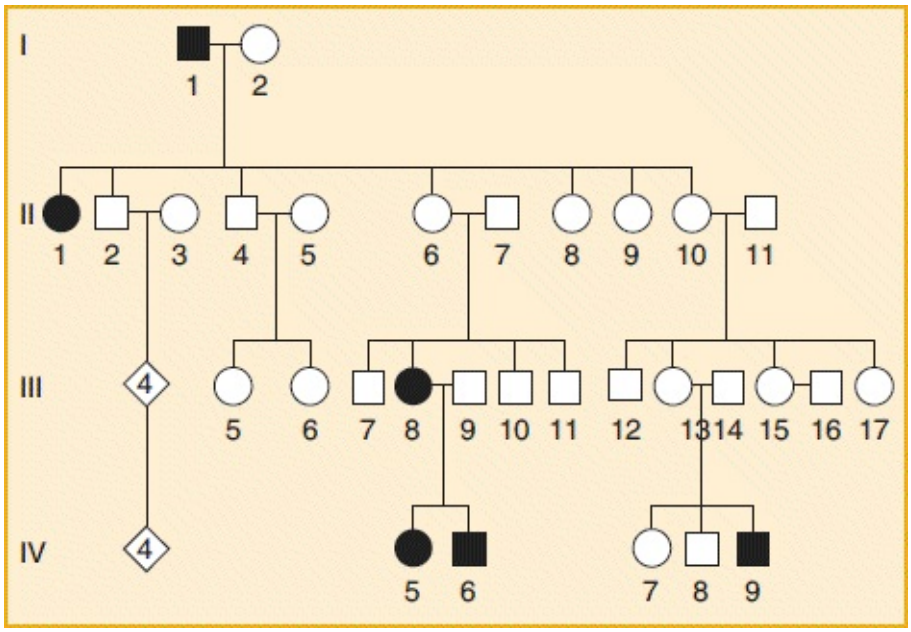
epistasia
expressividade
gene essencial
heterocário
hipótese um gene–um polipeptídio
letal sintético
modificador
mutação dominante negativa
mutação nula
mutantes duplos
mutações sensíveis à temperatura (st)
penetrância
revertente
RNA funcional
série alélica (alelos múltiplos)
supressor
temperatura permissiva
temperatura restritiva
teste de complementação

PROBLEMAS RESOLVIDOS

Problema resolvido 1. A maior parte dos heredogramas demonstra a polidactilia (ver Figura 2.25) herdada como autossômica dominante rara, mas os heredogramas de algumas famílias não se ajustam totalmente aos padrões esperados em relação à referida herança. Um referido heredograma está demonstrado aqui. (Os losangos não sombreados fazem referência aos números especificados de pessoas não afetadas de sexo desconhecido.)

a. Qual irregularidade esse heredograma demonstra?

b. Qual fenômeno genético esse heredograma ilustra?



c. Sugira um mecanismo de interação gênica específico que possa produzir um referido heredograma demonstrando os genótipos dos familiares pertinentes.

Solução

a. A expectativa normal em relação a um autossômico dominante é que cada indivíduo afetado apresente um progenitor afetado, mas essa expectativa não é observada nesse heredograma, o que constitui a irregularidade. Quais são algumas possíveis explicações?

Alguns casos de polidactilia poderiam ser causados por um gene diferente, um que seja um gene dominante ligado ao X? Essa sugestão não é útil, tendo em vista que ainda precisamos explicar a ausência da condição nas pessoas II-6 e II-10. Além disso, postular a herança recessiva, seja autossômica ou ligada ao sexo, exige que muitas pessoas no heredograma sejam heterozigotas, o que é inadequado, tendo em vista que a polidactilia é uma condição rara.

b. Portanto, resta-nos a conclusão de que a polidactilia por vezes é incompletamente penetrante. Conforme descrito neste capítulo, alguns indivíduos que apresentam o genótipo em relação a um fenótipo em particular não o expressam. Nesse heredograma, II-6 e II-10 aparentam pertencer a essa categoria; eles obrigatoriamente carreiam o gene da polidactilia herdado de I-1, tendo em

vista que o transmitem à sua progênie.

c. Conforme discutido neste capítulo, a supressão ambiental da expressão gênica pode causar penetrância incompleta, assim como a supressão por outro gene. Para fornecer a explicação genética solicitada, é preciso elaborar uma hipótese genética. O que precisamos explicar? A chave é que I-1 transmite a mutação para dois tipos de progênie, representadas por II-1, que expressa o fenótipo mutante, e por II-6 e II-10, que não expressam. (A partir do heredograma, não podemos dizer se os outros filhos de I-1 apresentam o alelo mutante.) A supressão genética está atuando? I-1 não apresenta um alelo supressor, tendo em vista que expressa polidactilia. Assim, a única pessoa da qual o supressor poderia advir é I-2. Além disso, I-2 deve ser heterozigoto em relação ao alelo supressor, tendo em vista que no mínimo um de seus filhos expressa polidactilia. Portanto, o alelo supressor tem de ser dominante. Sendo assim, formulamos a hipótese de que o cruzamento na geração I deve ter sido:

$$\text{(I-1) } P/p \cdot s/s \times \text{(I-2) } p/p \cdot S/s,$$

em que *S* é o supressor e *P* é o alelo responsável pela polidactilia. A partir dessa hipótese, prevemos que a progênie será composta pelos quatro tipos a seguir, se os genes se distribuírem:

Genótipo	Fenótipo	Exemplo
$P/p \cdot S/s$	normal (suprimido)	II-6, II-10
$P/p \cdot s/s$	polidáctilo	II-1
$p/p \cdot S/s$	normal	
$p/p \cdot s/s$	normal	

Se *S* for raro, as progênies de II-6 e II-10 serão:

Genótipo da progênie	Exemplo
P/p · S/s	III-13
P/p · s/s	III-8
p/p · S/s	
p/p · s/s	

Não podemos descartar as possibilidades de que II-2 e II-4 apresentem o genótipo *P/p · S/s* e que, ao acaso, nenhum de seus descendentes seja afetado.

Problema resolvido 2. Besouros de uma determinada espécie podem apresentar coberturas das asas verde, azul ou turquesa. Besouros virgens foram selecionados a partir de uma população de laboratório polimórfica e foram cruzados para determinar a herança da cor das coberturas das asas. Os cruzamentos e os resultados foram conforme fornecidos na tabela a seguir:

a. Deduza a base genética da cor das coberturas das asas nessa espécie.

b. Escreva os genótipos de todos os genitores e da progênie tão completamente quanto possível.

Cruzamento	Genitores	Progênie
1	Azul × Verde	toda azul
2	Azul × Azul	$\frac{3}{4}$ azul:$\frac{1}{4}$ turquesa
3	Verde × Verde	$\frac{3}{4}$ verde:$\frac{1}{4}$ turquesa
4	Azul × Turquesa	$\frac{1}{2}$ azul:$\frac{1}{2}$ turquesa
5	Azul × Azul	$\frac{3}{4}$ azul:$\frac{1}{4}$ verde
6	Azul × Verde	$\frac{1}{2}$ azul:$\frac{1}{2}$ verde
7	Azul × Verde	$\frac{1}{2}$ azul:$\frac{1}{4}$ verde $\frac{1}{4}$ turquesa
8	Turquesa × Turquesa	toda turquesa

Solução

a. Esses dados primeiramente aparentam ser complexos, mas o padrão de herança se torna claro se consideramos cada um dos cruzamentos por vez. Um princípio geral para a solução dos referidos problemas, conforme vimos, é iniciar observando todos os cruzamentos e agrupando os dados para revelar os padrões.

Um indício que surge de uma visão geral dos dados é que todas as proporções são proporções monogênicas: absolutamente não existe evidência da participação de dois genes em separado. Como essa variação pode ser explicada com um gene único? A resposta é que existe variação em relação ao próprio gene único — ou seja, alelismo múltiplo. Talvez existam três alelos de um gene; denominaremos o gene *w* (em relação à cor das coberturas das asas) e representaremos os alelos como w^g, w^b e w^t. Agora temos um problema adicional, que é determinar a dominância desses alelos.

O cruzamento 1 nos informa algo a respeito da dominância, tendo em vista que toda a progênie de um cruzamento azul × verde é azul; portanto, o azul aparenta ser dominante sobre o verde. Essa conclusão é amparada pelo cruzamento 5, tendo em vista que o determinante verde obrigatoriamente existia no estoque parental para aparecer na progênie. O cruzamento 3 nos informa a respeito dos determinantes de turquesa, que obrigatoriamente existiam no estoque parental, embora não fossem expressos, tendo em vista que existem coberturas das asas

turquesa na progênie. Assim, verde tem de ser dominante sobre turquesa. Portanto, formamos um modelo no qual a dominância é $w^b > w^g > w^t$. De fato, a posição inferida do alelo w^t na parte inferior da série de dominância é amparada pelos resultados do cruzamento 7, no qual turquesa aparece na progênie de um cruzamento azul × verde.

b. Agora se trata apenas de uma questão de deduzir os genótipos específicos. Observe que a questão declara que os genitores foram obtidos de uma população polimórfica, o que significa que eles podem ser homozigotos ou heterozigotos. Um genitor com coberturas das asas azuis, por exemplo, pode ser homozigoto (w^b/w^b) ou heterozigoto (w^b/w^g ou w^b/w^t). Aqui, são necessários um pouco de tentativa e erro e bom senso, mas, nesse estágio, a questão foi essencialmente respondida, e tudo o que resta é "cortar os t e pôr os pontos nos i". Os genótipos a seguir explicam os resultados. Um travessão indica que o genótipo pode ser homozigoto ou heterozigoto na apresentação de um segundo alelo na série alélica.

Cruzamento	Genitores	Progênie
1	$w^b/w^b \times w^g/-$	w^b/w^g ou $w^b/-$
2	$w^b/w^t \times w^b/w^t$	$\frac{3}{4}\ w^b/-:\frac{1}{4}\ w^t/w^t$
3	$w^g/w^t \times w^g/w^t$	$\frac{3}{4}\ w^g/-:\frac{1}{4}\ w^t/w^t$
4	$w^b/w^t \times w^t/w^t$	$\frac{1}{2}\ w^b/w^t:\frac{1}{2}\ w^t/w^t$
5	$w^b/w^g \times w^b/w^g$	$\frac{3}{4}\ w^b/-:\frac{1}{4}\ w^g/w^g$
6	$w^b/w^g \times w^g/w^g$	$\frac{1}{2}\ w^b/w^g:\frac{1}{2}\ w^g/w^g$
7	$w^b/w^t \times w^g/w^t$	$\frac{1}{2}\ w^b/-:\frac{1}{4}\ w^g/w^t:\frac{1}{4}\ w^t/w^t$
8	$w^t/w^t \times w^t/w^t$	*todas* w^t/w^t

Problema resolvido 3. As folhas de abacaxi podem ser classificadas em três tipos: espinhosas (S), com pontas espinhosas (ST), e tubulares (não espinhosas; P). Em cruzamentos entre linhagens puras seguidos por intercruzamentos da F_1, apareceram os resultados a seguir:

Fenótipos

Cruzamento	Parental	F_1	F_2
1	ST × S	ST	99 ST:34 S
2	P × ST	P	120 P:39 ST
3	P × S	P	95 P:25 ST:8 S

a. Atribua símbolos gênicos. Explique esses resultados em relação aos genótipos produzidos e suas proporções.
b. Com a utilização do modelo da parte *a*, forneça as proporções fenotípicas que você esperaria se cruzasse (1) a progênie F_1 de tubulares × espinhosas com o estoque parental espinhoso e (2) a progênie F_1 de tubulares × espinhosas com a progênie F_1 de espinhosas × pontas espinhosas.

Solução

a. Primeiramente, vejamos as proporções da F_2. Temos claras proporções de 3:1 nos cruzamentos 1 e 2, indicando segregações monogênicas. Entretanto, o cruzamento 3 demonstra uma proporção que é quase certamente uma proporção de 12:3:1. Como conhecemos essa proporção? Bem, simplesmente não existem tantas proporções complexas em genética, e a tentativa e erro nos traz à proporção de 12:3:1 de modo consideravelmente rápido. No total de 128 da progênie, são esperados os números 96:24:8, mas os números reais correspondem notavelmente bem a essas expectativas.

Um dos princípios deste capítulo é que as proporções mendelianas modificadas revelam interações gênicas. O cruzamento 3 proporciona números de F_2 apropriados para uma proporção mendeliana di-híbrida modificada e, assim, aparentemente estamos lidando com uma interação de dois genes. Esse aparenta ser o lugar mais promissor para o início; podemos retornar aos cruzamentos 1 e 2 e tentar ajustá-los posteriormente.

Qualquer proporção di-híbrida tem por base as proporções fenotípicas de

9:3:3:1. Nossa modificação observada as agrupa como segue:

9 *A*/—;*B*/—	12 tubulares
3 *A*/—;*b*/*b*	
3 *a*/*a*;*B*/—	3 pontas espinhosas
1 *a*/*a*;*b*/*b*	1 espinhosa

Assim, sem nos preocuparmos a respeito da denominação do tipo de interação gênica (de qualquer maneira, não nos é solicitado o seu fornecimento), já podemos definir nossos três fenótipos das folhas do abacaxi em relação aos pares alélicos propostos *A*/*a* e *B*/*b*:

Tubular = *A*/— (*B*/*b* irrelevante)
Ponta espinhosa = *a*/*a*; *B*/—
Espinhosa = *a*/*a*; *b*/*b*

O que dizer sobre os genitores do cruzamento 3? O genitor espinhoso tem de ser *a*/*a*; *b*/*b* e, tendo em vista que o gene *B* é necessário para produzir folhas com pontas espinhosas na F_2, o genitor tubular tem de ser *A*/*A*; *B*/*B*. (Observe que nos é informado que todos os genitores são puros, ou homozigotos.) Portanto, a F_1 deve ser *A*/*a*; *B*/*b*.

Sem pensamentos adicionais, podemos escrever nosso cruzamento 1 como segue:

$$a/a\ ;\ B/B\ \times\ a/a\ ;\ b/b\ \longrightarrow\ a/a\ ;\ B/b \begin{cases} \nearrow \frac{3}{4}\ a/a\ ;B/- \\ \searrow \frac{1}{4}\ a/a\ ;\ b/b \end{cases}$$

O cruzamento 2 pode ser escrito parcialmente sem pensamentos adicionais utilizando nossos símbolos gênicos arbitrários:

$$A/A\ ;\ -/-\ \times\ a/a\ ;\ B/B\ \longrightarrow\ A/a\ ;\ B/- \begin{cases} \nearrow \frac{3}{4}\ A/-\ ;\ -/- \\ \searrow \frac{1}{4}\ a/a\ ;\ B/- \end{cases}$$

Sabemos que a F_2 do cruzamento 2 demonstra segregação monogênica, e agora aparenta estar certo que há participação do par alélico *A/a*. Mas o alelo *B* é necessário para produzir o fenótipo ponta espinhosa e, assim, todas as plantas devem ser homozigotas *B/B*:

$$A/A\ ;\ B/B\ \times\ a/a\ ;\ B/B\ \longrightarrow\ A/a\ ;\ B/B \begin{cases} \nearrow \frac{3}{4}\ A/-\ ;\ B/B \\ \searrow \frac{1}{4}\ a/a\ ;\ B/B \end{cases}$$

Observe que as duas segregações monogênicas nos cruzamentos 1 e 2 não demonstram que os genes *não* estão interagindo. O que está demonstrado é que a interação de dois genes não é *revelada* por esses cruzamentos — apenas pelo cruzamento 3, no qual a F_1 é heterozigota em relação a ambos os genes.

b. Agora é simplesmente uma questão de utilizar as leis de Mendel para prever os desfechos dos cruzamentos:

(1) $A/a\ ;\ B/b\ \times\ a/a\ ;\ b/b\ \longrightarrow$

(distribuição independente em um cruzamento-teste padrão)

- $\frac{1}{4}\ A/a\ ;\ B/b$ — tubular
- $\frac{1}{4}\ A/a\ ;\ b/b$ — tubular
- $\frac{1}{4}\ a/a\ ;\ B/b$ — ponta espinhosa
- $\frac{1}{4}\ a/a\ ;\ b/b$ — espinhosa

(2) $A/a\ ;\ B/b\ \times\ a/a\ ;\ B/b\ \longrightarrow$

- $\frac{1}{2}\ A/a$ → $\frac{3}{4}\ B/- \longrightarrow \frac{3}{8}$ — $\frac{1}{2}$ tubular
- $\frac{1}{2}\ A/a$ → $\frac{1}{4}\ b/b \longrightarrow \frac{1}{8}$ — $\frac{1}{2}$ tubular
- $\frac{1}{2}\ a/a$ → $\frac{3}{4}\ B/- \longrightarrow \frac{3}{8}$ — ponta espinhosa
- $\frac{1}{2}\ a/a$ → $\frac{1}{4}\ b/b \longrightarrow \frac{1}{8}$ — espinhosa

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 6.1:

a. o que as estrelas amarelas representam?

b. explique em suas próprias palavras por que o heterozigoto é funcionalmente do tipo selvagem.

2. Na Figura 6.2, explique como o polipeptídio mutante atua como um espoliador e qual é o seu efeito líquido sobre o fenótipo.

3. Na Figura 6.6, avalie o alelo V^{f} em relação ao alelo V^{by}: ele é dominante?; recessivo?; codominante?; incompletamente dominante?

4. Na Figura 6.11,

a. em virtude da posição da HPA oxidase anteriormente na via bioquímica em comparação àquela da HA oxidase, você espera que pessoas com tirosinemia demonstrem sintomas de alcaptonúria?

b. se um mutante duplo puder ser encontrado, você espera que a tirosinemia seja epistática em relação à alcaptonúria?

5. Na Figura 6.12,

a. o que os símbolos de dólar, libra e iene representam?

b. por que o heterozigoto da F_1 à esquerda não consegue sintetizar o pigmento azul?

6. Na Figura 6.13, explique no nível proteico por que esse heterocário consegue crescer em meio mínimo.

7. Na Figura 6.14, escreva os possíveis genótipos em relação a cada uma das quatro cobras ilustradas.

8. Na Figura 6.15,

a. qual painel representa o mutante duplo?

b. declare a função do gene regulador.

c. na situação no painel B, a proteína do gene da proteína ativa seria produzida?

9. Na Figura 6.16, se você realizasse a autofecundação de 10 plantas cor-de-rosa da F_2 diferentes, você esperaria observar quaisquer plantas com flores brancas na progênie? Quaisquer plantas com flores azuis?

10. Na Figura 6.19:

a. o que as cavilhas quadradas/triangulares e os orifícios representam?

b. a mutação supressora isoladamente é de fenótipo do tipo selvagem?

11. Na Figura 6.21, proponha uma explicação genética específica em relação ao indivíduo Q (forneça um genótipo possível, definindo os alelos).

PROBLEMAS BÁSICOS

12. Em seres humanos, a doença galactosemia causa retardo mental em uma idade precoce. A lactose (o açúcar do leite) é fragmentada em galactose e glicose. Normalmente, a galactose é adicionalmente degradada pela enzima galactose-1-fosfato uridiltransferase (GALT). Entretanto, em pacientes com galactosemia, a GALT é inativa, levando ao acúmulo de níveis altos de galactose, os quais, no cérebro, causam retardo mental. Como você proporcionaria uma cura secundária para a galactosemia? Você espera que o fenótipo dessa doença seja dominante ou recessivo?

13. Em seres humanos, a PKU (fenilcetonúria) é uma doença recessiva causada por uma insuficiência enzimática na etapa A na sequência de ações simplificada a seguir, e a AKU (alcaptonúria) é outra doença recessiva em virtude de uma ineficiência enzimática em uma das etapas resumida aqui como a etapa B:

$$\text{fenilalanina} \xrightarrow{A} \text{tirosina} \xrightarrow{B} CO_2 + H_2O$$

Uma pessoa com PKU se casa com uma pessoa com AKU. Quais fenótipos você espera em relação aos seus filhos? Todos normais, todos apresentando apenas PKU, todos apresentando apenas AKU, todos apresentando ambas, PKU e AKU, ou alguns apresentando AKU e alguns apresentando PKU?

14. Na *Drosophila*, o autossômico recessivo *bw* causa olhos marrom-escuros, e o autossômico recessivo não ligado *st* causa em olhos escarlate-vivo. Um homozigoto em relação a ambos os genes apresenta olhos brancos. Portanto, temos as correspondências a seguir entre os genótipos e os

fenótipos:

st^+/st^+; bw^+/bw^+ = Olho vermelho (tipo selvagem)
st^+/st^+; *bw/bw* = Olho marrom
st/st; bw^+/bw^+ = Olho escarlate
st/st; *bw/bw* = Olho branco

Construa uma via de biossíntese hipotética demonstrando como os produtos gênicos interagem e por que as diferentes combinações de mutantes apresentam fenótipos diferentes.

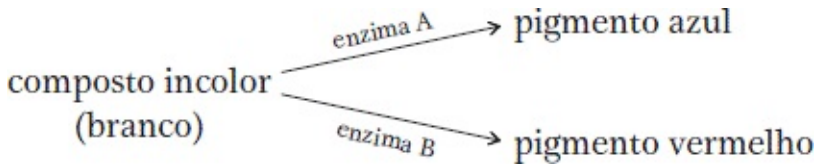
15. Diversos mutantes são isolados, todos os quais necessitam do composto G para o crescimento. Os compostos (A a E) na via de biossíntese para G são conhecidos, mas a sua ordem na via não é conhecida. Cada composto é testado em relação à sua capacidade de estimular o crescimento de cada mutante (1 a 5). Na tabela a seguir, um sinal de mais indica crescimento e um sinal de menos indica ausência de crescimento.

		Composto testado					
		A	B	C	D	E	G
Mutante	1	—	—	—	+	—	+
	2	—	+	—	+	—	+
	3	—	—	—	—	—	+
	4	—	+	+	+	—	+
	5	+	+	+	+	—	+

a. Qual é a ordem dos compostos A a E na via?
b. Em que ponto na via cada mutante é bloqueado?

c. Um heterocário composto por mutantes duplos 1,3 e 2,4 cresceria em um meio mínimo? 1,3 e 3,4 cresceriam? 1,2 e 2,4 e 1,4 cresceriam?

16. Em uma determinada planta, as pétalas das flores normalmente são roxas. Duas mutações recessivas surgem em plantas separadas e observa-se que estão em cromossomos diferentes. A mutação 1 (m_1) proporciona pétalas azuis quando em homozigose (m_1/m_1). A mutação 2 (m_2) proporciona pétalas vermelhas quando em homozigose (m_2/m_2). Bioquímicos que trabalham com a síntese dos pigmentos das flores nessa espécie já descreveram a via a seguir:



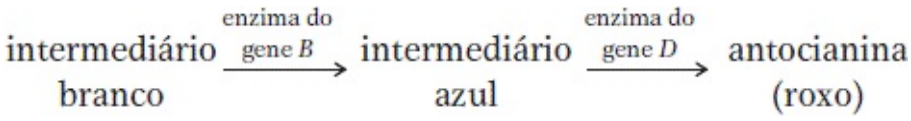
a. Qual mutante você espera que seja deficiente na atividade da enzima A?

b. Uma planta apresenta o genótipo M_1/m_1; M_2/m_2. Qual você espera que seja o seu fenótipo?

c. Se a planta na parte *b* for autofecundada, quais cores da progênie você espera e em quais proporções?

d. Por que esses mutantes são recessivos?

17. Em ervilhas-de-cheiro, a síntese do pigmento antocianina roxo nas pétalas é controlada por dois genes, *B* e *D*. A via é:



a. Quais cores de pétalas você espera em uma planta pura incapaz de catalisar a primeira reação?

b. Quais cores de pétalas você espera em uma planta pura incapaz de catalisar a segunda reação?

c. Se as plantas nas partes *a* e *b* forem cruzadas, quais cores de pétalas as plantas da F_1 apresentarão?

d. Qual proporção de plantas roxas:azuis:brancas você esperaria na F_2?

18. Se um homem do grupo sanguíneo AB se casar com uma mulher do grupo sanguíneo A cujo pai era do grupo sanguíneo O, a quais diferentes grupos sanguíneos esse homem e essa mulher podem esperar que seus filhos pertençam?

19. A maior parte das penas da ave *erminette* é de cor clara, com uma preta ocasional, proporcionando um aspecto pintado. Um cruzamento de duas *erminette* produziu uma progênie total de 48, composta por 22 *erminettes*, 14 pretas, e 12 brancas puras. Qual base genética do padrão *erminette* é sugerida? Como você testaria a sua hipótese?

20. Os rabanetes podem ser longos, redondos ou ovais, e podem ser vermelhos, brancos ou roxos. Você cruza uma variedade longa e branca com uma variedade redonda e vermelha e obtém uma F_1 oval e roxa. A F_2 demonstra nove classes fenotípicas, como segue: 9 longos e vermelhos; 15 longos e roxos; 19 ovais e vermelhos; 32 ovais e roxos; 8 longos e brancos; 16 redondos e roxos; 8 redondos e brancos; 16 ovais e brancos; e 9 redondos e vermelhos.

a. Forneça uma explicação genética desses resultados. Assegure-se de definir os genótipos e demonstrar a constituição dos genitores, da F_1 e da F_2.

b. Preveja as proporções genotípicas e fenotípicas na progênie de um cruzamento entre um rabanete longo e roxo e um rabanete oval e roxo.

21. Na série de alelos múltiplos que determina a cor da pelagem em coelhos, c^+ codifica aguti, c^{ch} codifica chinchila (uma cor de pelagem bege), e c^h codifica himalaio. A dominância é na ordem $c^+ > c^{ch} > c^h$. Em um cruzamento de $c^+/c^{ch} \times c^{ch}/c^h$, qual proporção da progênie será chinchila?

22. Preto, sépia, creme e albino são cores da pelagem de cobaias. Animais individuais (não necessariamente de linhagens puras) que demonstram essas cores foram entrecruzados; os resultados estão tabulados como segue, em que as abreviações A (albino), B (preto), C (creme) e S (sépia) representam os fenótipos:

Cruzamento	Fenótipos parentais	Fenótipos da progênie			
		B	S	C	A
1	B × B	22	0	0	7
2	B × A	10	9	0	0
3	C × C	0	0	34	11
4	S × C	0	24	11	12
5	B × A	13	0	12	0
6	B × C	19	20	0	0
7	B × S	18	20	0	0
8	B × S	14	8	6	0
9	S × S	0	26	9	0
10	C × A	0	0	15	17

a. Deduza a herança dessas cores de pelagem, e utilize símbolos gênicos de sua própria escolha. Demonstre todos os genótipos dos genitores e da progênie.

b. Se os animais pretos nos cruzamentos 7 e 8 forem cruzados, quais proporções de progênie você pode prever ao utilizar o seu modelo?

23. Em uma ala de maternidade, quatro recém-nascidos são misturados acidentalmente. Os tipos AB0 dos quatro bebês sabidamente são 0, A, B e AB. Os tipos AB0 dos quatro conjuntos de pais são determinados. Indique

qual recém-nascido pertence a cada conjunto de pais: **(a)** AB × 0, **(b)** A × 0, **(c)** A × AB, **(d)** 0 × 0.

24. Considere dois polimorfismos sanguíneos que os seres humanos apresentam adicionalmente ao sistema AB0. Dois alelos L^M e L^N determinam os grupos sanguíneos M, N e MN. O alelo dominante *R* de um gene diferente faz com que uma pessoa apresente o fenótipo Rh^+ (*rhesus* positivo), enquanto o homozigoto em relação a *r* é Rh^- (*rhesus* negativo). Dois homens levaram uma disputa de paternidade aos tribunais, cada um reclamando que três crianças são suas. Os grupos sanguíneos dos homens, das crianças e de sua mãe eram como segue:

Pessoa	Grupo sanguíneo		
marido	0	M	Rh^+
amante da esposa	AB	MN	Rh^-
esposa	A	N	Rh^+
criança 1	0	MN	Rh^+
criança 2	A	N	Rh^+
criança 3	A	MN	Rh^-

A partir dessas evidências, a paternidade das crianças pode ser estabelecida?

25. Em uma fazenda de raposas em Wisconsin, surgiu uma mutação que proporcionou uma cor de pelagem "platina". A cor platina comprovou ser muito popular entre compradores de casacos de raposa, mas os criadores não conseguiram desenvolver uma linhagem platina pura. Cada vez que

duas platinas eram cruzadas, apareciam algumas raposas normais na progênie. Por exemplo, os cruzamentos repetidos do mesmo par de platinas produziram progênie de 82 platinas e 38 normais. Todos os outros referidos cruzamentos forneceram proporções de progênie semelhantes. Declare uma hipótese genética concisa que explique esses resultados.

26. Durante diversos anos, Hans Nachtsheim investigou uma anomalia hereditária de leucócitos de coelhos. Essa anomalia, denominada *anomalia de Pelger*, é a parada da segmentação dos núcleos de determinados leucócitos. Essa anomalia não aparenta afetar seriamente os coelhos.

a. Quando coelhos que demonstram a anomalia de Pelger foram cruzados com coelhos de um estoque puro normal, Nachtsheim contou 217 descendentes demonstrando a anomalia de Pelger e 237 descendentes normais. Qual é a base genética da anomalia de Pelger?

b. Quando coelhos com a anomalia de Pelger foram cruzados entre si, Nachtsheim observou na progênie 223 normais, 439 com a anomalia de Pelger e 39 extremamente anormais. Essa progênie muito anormal não apenas apresentava leucócitos defeituosos, mas também demonstrava deformidades graves do sistema esquelético; quase toda ela morreu logo após o nascimento. Em termos genéticos, o que você supõe que esses coelhos extremamente defeituosos tenham representado? Por que existem apenas 39 deles?

c. Qual evidência experimental adicional você poderia coletar para testar a sua hipótese na parte *b*?

d. Em Berlim, aproximadamente 1 ser humano em 1.000 demonstra uma anomalia de Pelger de leucócitos muito semelhante àquela descrita em relação aos coelhos. A anomalia é herdada como um dominante simples, mas o tipo homozigoto não foi observado em humanos. Com base na condição em coelhos, por que você supõe que o homozigoto humano não foi observado?

e. Novamente por meio da analogia com os coelhos, quais fenótipos e genótipos você esperaria entre os filhos de um homem e uma mulher que

demonstram, ambos, a anomalia de Pelger?

(Dados de A. M. Srb, R. D. Owen and R. S. Edgar, *General Genetics*, 2nd ed. W. H. Freeman and Company, 1965.)

27. Duas moscas-das-frutas de aspecto normal foram cruzadas e, na progênie, havia 202 fêmeas e 98 machos.

a. O que é incomum a respeito desse resultado?

b. Forneça uma explicação genética para essa anomalia.

c. Forneça um teste para sua hipótese.

28. Você recebeu uma fêmea de *Drosophila* virgem. Você observa que as cerdas em seu tórax são muito mais curtas do que o normal. Você a cruza com um macho normal (com cerdas longas) e obtém a progênie de F_1 a seguir: $\frac{1}{3}$ de fêmeas com cerdas curtas, $\frac{1}{3}$ de fêmeas com cerdas longas, e $\frac{1}{3}$ de machos com cerdas longas. Um cruzamento das fêmeas com cerdas longas da F_1 com seus irmãos fornece apenas uma F_2 com cerdas longas. Um cruzamento de fêmeas com cerdas curtas com seus irmãos fornece $\frac{1}{3}$ de fêmeas com cerdas curtas, $\frac{1}{3}$ de fêmeas com cerdas longas e $\frac{1}{3}$ de machos com cerdas longas. Forneça uma hipótese genética para explicar todos esses resultados, demonstrando os genótipos em cada cruzamento.

29. Um alelo dominante *H* reduz a quantidade de cerdas corporais que as moscas *Drosophila* apresentam, dando origem a um fenótipo "calvo". Na condição homozigota, *H* é letal. Um alelo dominante de distribuição independente *S* não apresenta efeitos sobre a quantidade de cerdas, exceto na presença de *H*, caso em que uma dose única de *S* suprime o fenótipo calvo, restaurando, assim, o fenótipo com cerdas. Entretanto, *S* também é letal na condição homozigota (*S/S*).

a. Qual proporção de moscas com e sem cerdas você encontraria na progênie viva de um cruzamento entre duas moscas com cerdas carreadoras de *H* na condição suprimida?

b. Quando a progênie sem cerdas é retrocruzada com uma mosca parental com cerdas, qual proporção fenotípica você esperaria observar na sua progênie viva?

30. Após irradiar células do tipo selvagem de *Neurospora* (um fungo haploide), um geneticista observa dois mutantes auxotróficos que necessitam de leucina. Ele combina os dois mutantes em um heterocário e descobre que o heterocário é prototrófico.

a. As mutações nos dois auxotróficos estão no *mesmo* gene na via para a síntese de leucina ou em dois genes *diferentes* naquela via? Explique.

b. Escreva o genótipo das duas linhagens de acordo com o seu modelo.

c. Qual progênie e em quais proporções você preveria a partir do cruzamento dos dois mutantes auxotróficos? (Presuma a distribuição independente.)

31. Um geneticista de leveduras irradia células haploides de uma linhagem que é um mutante auxotrófico que necessita de adenina, causada pela mutação do gene *ade1*. Milhões de células irradiadas são plaqueadas em meio mínimo, e um pequeno número de células se divide e produz colônias prototróficas. Essas colônias são cruzadas individualmente com uma linhagem do tipo selvagem. Dois tipos de resultados são obtidos:

(1) Prototrófica × Tipo selvagem: todas a progênie prototrófica

(2) Prototrófica × Tipo selvagem: progênie 75% prototrófica, 25% auxotrófica que necessita de adenina

a. Explique a diferença entre esses dois tipos de resultados.

b. Escreva os genótipos dos prototróficos em cada caso.

c. Quais fenótipos e proporções da progênie você prevê a partir do cruzamento de um prototrófico do tipo 2 com o auxotrófico *ade1* original?

32. Em rosas, a síntese do pigmento vermelho ocorre por meio de duas etapas em uma via, como segue:

$$\text{intermediário incolor} \xrightarrow{\text{gene } P} \text{intermediário magenta} \xrightarrow{\text{gene } Q} \text{pigmento vermelho}$$

a. Qual seria o fenótipo de uma planta homozigota em relação a uma mutação nula do gene *P*?

b. Qual seria o fenótipo de uma planta homozigota em relação a uma mutação nula do gene *Q*?

c. Qual seria o fenótipo de uma planta homozigota em relação a mutações nulas dos genes *P* e *Q*?

d. Escreva os genótipos das três linhagens nas partes *a*, *b* e *c*.

e. Qual proporção da F_2 é esperada a partir do cruzamento de plantas das partes *a* e *b*? (Presuma a distribuição independente.)

33. Tendo em vista que bocas-de-leão (*Antirrhinum*) possuem o pigmento antocianina, elas apresentam pétalas roxas avermelhadas. Foram desenvolvidas duas linhagens sem antocianina puras de *Antirrhinum*, uma na Califórnia e uma na Holanda. Elas aparentavam ser idênticas, sem apresentar pigmento vermelho em absoluto, manifestadas como flores brancas (albinas). Entretanto, quando pétalas das duas linhagens foram trituradas juntas em tampão no mesmo tubo de ensaio, a solução, que primeiramente aparentava ser incolor, gradualmente se tornou vermelha.

a. Quais experimentos de controle um investigador deve conduzir antes de proceder com a análise adicional?

b. O que poderia explicar a produção da cor vermelha no tubo de ensaio?

c. De acordo com a sua explicação em relação à parte *b*, quais seriam os genótipos das duas linhagens?

d. Se as duas linhagens brancas fossem cruzadas, quais você esperaria que fossem os fenótipos da F_1 e da F_2?

34. A ave *frizzle* é muito admirada por apreciadores de galinhas. Ela obtém sua denominação em virtude do modo incomum como as suas penas se encurvam, dando a impressão de que foram (nas palavras memoráveis do geneticista animal F. B. Hutt) “puxadas para trás através de um nó da madeira”. Infelizmente, *frizzle* não é pura: quando duas *frizzle* são entrecruzadas, elas sempre produzem 50% de aves *frizzle*, 25% de normais e 25% com penas lanosas peculiares, que logo em seguida caem, deixando as aves sem plumagem.

a. Forneça uma explicação genética para esses resultados, demonstrando os

genótipos de todos os fenótipos, e forneça uma declaração de como a sua explicação funciona.

b. Se você quisesse produzir aves *frizzle* em massa para a venda, quais tipos seriam os melhores para serem utilizados como um par de cruzamento?

35. As pétalas da planta *Collinsia parviflora* normalmente são azuis. Duas linhagens puras foram obtidas a partir de variantes coloridas encontradas na natureza; a primeira linhagem apresentava pétalas rosa, e a segunda linhagem apresentava pétalas brancas. Foram realizados os cruzamentos a seguir entre as linhagens puras, com os resultados demonstrados:

Genitores	F_1	F_2
Azul × Branca	azul	101 azuis, 33 brancas
Azul × Rosa	azul	192 azuis, 63 rosa
Rosa × Branca	azul	272 azuis, 121 brancas, 89 rosa

a. Explique esses resultados geneticamente. Defina os símbolos alélicos que você utilizar e demonstre a constituição genética dos genitores, da F_1 e da F_2 em cada cruzamento.

b. Um cruzamento entre uma determinada planta da F_2 azul e uma determinada planta da F_2 branca forneceu progênie, da qual $\frac{3}{8}$ eram azuis, $\frac{1}{8}$ eram rosa e $\frac{1}{2}$ eram brancas. Quais devem ter sido os genótipos dessas duas plantas da F_2?

Como solucionar problema 35

1. Qual é a característica que está sendo estudada?
2. Qual é o fenótipo do tipo selvagem?

3. O que é uma variante?
4. Quais são as variantes neste problema?
5. O que significa “na natureza”?
6. De qual modo as variantes teriam sido encontradas na natureza? (Descreva a cena.)
7. Em quais estágios nos experimentos seriam utilizadas sementes?
8. O modo de escrever um cruzamento “azul × branca”, por exemplo, significaria o mesmo que “branca × azul”? Você esperaria resultados semelhantes? Por que sim ou por que não?
9. De que modo as duas primeiras fileiras na tabela diferem da terceira fileira?
10. Quais fenótipos são dominantes?
11. O que é complementação?
12. De onde advém a característica azul na progênie do cruzamento rosa × branca?
13. Qual fenômeno genético a produção de uma F_1 azul de genitores rosa e branco representa?
14. Liste quaisquer proporções que você consiga visualizar.
15. Existem quaisquer proporções mono-híbridas?
16. Existem quaisquer proporções di-híbridas?
17. O que a observação das proporções mono-híbridas e di-híbridas informa a você?
18. Liste quatro proporções mendelianas modificadas a respeito das quais você consiga pensar.
19. Existem quaisquer proporções mendelianas modificadas no problema?
20. O que as proporções mendelianas modificadas em geral indicam?
21. O que é indicado pela proporção ou pelas proporções modificadas específicas neste problema?

22. Desenhe os cromossomos representando as meioses nos genitores no cruzamento azul × branca e representando a meiose na F_1.

23. Repita a etapa 22 em relação ao cruzamento azul × rosa.

36. Uma mulher que possuía um poodle albino puro (um fenótipo autossômico recessivo) desejava filhotes brancos; assim, ela levou o cão até um criador, que lhe disse que cruzaria a fêmea com um macho reprodutor albino, também de um estoque puro. Quando nasceram seis filhotes, todos eles eram pretos; assim, a mulher processou o criador, reclamando que ele substituiu o macho reprodutor por um cão preto, dando a ela seis filhotes indesejados. Você é chamado como uma testemunha especialista, e a defesa indaga a você se é possível produzir descendência preta a partir de dois genitores albinos recessivos puros. Que testemunho você fornece?

37. Uma planta boca-de-leão pura para pétalas brancas foi cruzada com uma planta que produziu o mesmo fenótipo em relação a pétalas roxas e toda a F_1 apresentou pétalas brancas. A F_1 foi autofecundada. Entre a F_2, foram observados três fenótipos, nas quantidades a seguir:

branca	240
roxa sólida	61
roxa manchada	19
Total	320

a. Proponha uma explicação para esses resultados, demonstrando os genótipos de todas as gerações (crie e explique os seus símbolos).

b. Uma planta da F_2 branca foi cruzada com uma planta da F_2 roxa sólida, e a progênie foi:

Branca	50%
Roxa sólida	25%
Roxa manchada	25%

Quais eram os genótipos das plantas da F_2 cruzadas?

38. A maior parte dos besouros *Tribolium castaneum* é preta, mas são conhecidas diversas variantes coloridas. Cruzamentos de genitores puros produziram os resultados a seguir (ver tabela) na geração F_1, e o intercruzamento da F_1 a partir de cada cruzamento forneceu as proporções demonstradas em relação à geração F_2. Os fenótipos são abreviados Bl = Preto; Br = Marrom; Y = Amarelo; e W = Branco.

Cruzamento	Genitores	F_1	F_2
1	Br × Y	Br	3 Br:1 Y
2	Bl × Br	Bl	3 Bl:1 Br
3	Bl × Y	Bl	3 Bl:1 Y
4	W × Y	Bl	9 Bl:3 Y:4 W
5	W × Br	Bl	9 Bl:3 Br:4 W
6	Bl × W	Bl	9 Bl:3 Y:4 W

a. A partir desses resultados, deduza e explique a herança dessas cores.

b. Escreva os genótipos de cada um dos genitores, da F_1 e da F_2 em todos os cruzamentos.

39. Dois albinos se casam e têm quatro filhos normais. Como isso é possível?

40. Considere a produção da cor das flores na glória-da-manhã japonesa (*Pharbitis nil*). Os alelos dominantes de qualquer um de dois genes separados (*A/— · b/b* ou *a/a · B/—*) produzem pétalas roxas. *A/— · B/—* produz pétalas azuis, e *a/a · b/b* produz pétalas escarlate. Deduza os genótipos dos genitores e da progênie nos cruzamentos a seguir:

Cruzamento	Genitores	Progênie
1	Azul × Escarlate	$\frac{1}{4}$ azul:$\frac{1}{2}$ roxa:$\frac{1}{4}$ escarlate
2	Roxa × Roxa	$\frac{1}{4}$ azul:$\frac{1}{2}$ roxa:$\frac{1}{4}$ escarlate
3	Azul × Azul	$\frac{3}{4}$ azul:$\frac{1}{4}$ roxa
4	Azul × Roxa	$\frac{3}{8}$ azul:$\frac{4}{8}$ roxa:$\frac{1}{8}$ escarlate
5	Roxa × Escarlate	$\frac{1}{2}$ roxa:$\frac{1}{2}$ escarlate

41. Cultivadores de milho obtiveram linhagens puras, cujas espigas se tornam vermelho-sol, rosa, escarlate ou laranja quando expostas à luz solar (as espigas normais permanecem amarelas na luz solar). Alguns cruzamentos entre essas linhagens produziram os resultados a seguir. Os fenótipos são abreviados O = Laranja; P = Rosa; Sc = Escarlate; e SR = Vermelho-sol.

Cruzamento	Fenótipos		
	Genitores	F_1	F_2
1	SR × P	todos SR	66 SR:20 P
2	O × SR	todos SR	998 SR:314 O
3	O × P	todos O	1.300 O:429 P
4	O × Sc	todos Y	182 Y:80 O: 58 Sc

Analise os resultados de cada cruzamento e forneça uma hipótese unificadora para explicar *todos* os resultados. (Explique todos os símbolos que você utilizar.)

42. Muitos tipos de animais selvagens apresentam o padrão de cor aguti, no qual cada pelo apresenta uma faixa amarela ao seu redor.

a. Camundongos pretos e outros animais pretos não apresentam a faixa amarela; cada um de seus pelos é todo preto. Essa ausência do padrão aguti selvagem é denominada *não aguti*. Quando camundongos de uma linhagem aguti pura são cruzados com não agutis, a F_1 é toda aguti e a F_2 apresenta uma proporção de 3:1 de agutis e não agutis. Diagrame esse cruzamento, fazendo *A* representar o alelo responsável pelo fenótipo aguti e *a*, não aguti. Demonstre os fenótipos e os genótipos dos genitores, de seus gametas, da F_1, de seus gametas e da F_2.

b. Outro desvio da cor herdado em camundongos substitui a cor marrom pela preta no pelo do tipo selvagem. Os referidos camundongos aguti marrons são denominados *canela*. Quando camundongos do tipo selvagem são cruzados com canelas, toda a F_1 é do tipo selvagem e a F_2 apresenta uma proporção de 3:1 do tipo selvagem e canela. Diagrame esse cruzamento como na parte *a*, tendo *B* como referência ao alelo preto do tipo selvagem e *b* como referência ao alelo marrom canela.

c. Quando camundongos de uma linhagem canela pura são cruzados com camundongos de uma linhagem pura não aguti (preta), toda a F_1 é do tipo selvagem. Utilize um diagrama genético para explicar esse resultado.

d. Na F_2 do cruzamento na parte *c*, uma quarta cor, denominada *chocolate*, aparece além da canela e não aguti parental e do tipo selvagem da F_1. Camundongos chocolate apresentam uma cor marrom sólida e rica. Qual é a constituição genética deles?

e. Presumindo que os pares alélicos *A/a* e *B/b* distribuam-se independentemente entre si, quais você espera serem as frequências relativas dos quatro tipos de cores na F_2 descrita na parte *d*? Diagrame o cruzamento das partes *c* e *d*, demonstrando os fenótipos e os genótipos (incluindo os gametas).

f. Quais fenótipos seriam observados e em quais proporções na progênie de um retrocruzamento de camundongos da F_1 da parte *c* com o estoque parental canela? Com o estoque parental não aguti (preto)? Diagrame esses

retrocruzamentos.

g. Diagrame um cruzamento-teste em relação à F_1 da parte *c*. Quais cores resultariam e em quais proporções?

h. Camundongos albinos (brancos com olhos rosa) são homozigotos em relação ao membro recessivo de um par alélico *C/c*, que se distribui independentemente dos pares *A/a* e *B/b*. Suponha que você possui quatro diferentes linhagens albinas altamente endocruzadas (e, portanto, presumivelmente homozigotas). Você cruza cada uma dessas linhagens com uma linhagem do tipo selvagem pura, e obtém uma grande progênie F_2 a partir de cada cruzamento. Quais genótipos em relação às linhagens albinas você consegue deduzir a partir dos fenótipos da F_2 a seguir?

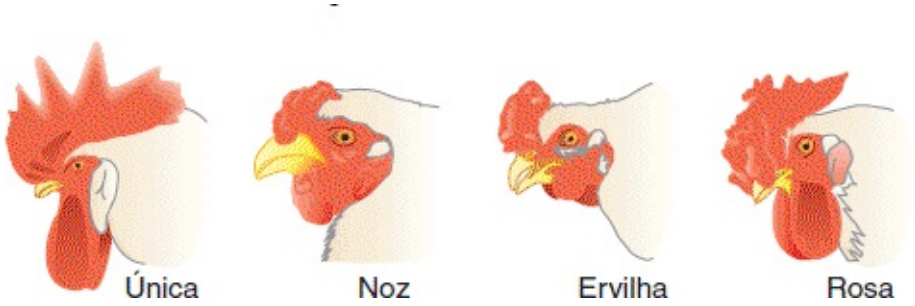
F_2 da linhagem	Fenótipos da progênie				
	Tipo selvagem	Preto	Canela	Chocolate	Albino
1	87	0	32	0	39
2	62	0	0	0	18
3	96	30	0	0	41
4	287	86	92	29	164

(Adaptada de A. M. Srb, R. D. Owen, and R. S. Edgar, *General Genetics*, 2nd ed. W. H. Freeman and Company, 1965.)

43. Um alelo *A* que não é letal quando homozigoto faz com que os ratos apresentem pelagem amarela. O alelo *R* de um gene em separado que se distribui independentemente produz pelagem preta. Em conjunto, *A* e *R* produzem pelagem acinzentada, enquanto *a* e *r* produzem pelagem branca. Um macho cinza é cruzado com uma fêmea amarela, e a F_1 é $\frac{3}{8}$ amarela, $\frac{3}{8}$

cinza, $\frac{1}{8}$ preta e $\frac{1}{8}$ branca. Determine os genótipos dos genitores.

44. O genótipo *r/r*; *p/p* proporciona às aves uma crista única, *R/—*; *P/—* proporciona uma crista em noz, *r/r*; *P/—* proporciona uma crista em ervilha, e *R/—*; *p/p* proporciona uma crista rosa (ver ilustrações). Presuma a distribuição independente.



a. Quais tipos de crista aparecerão na F_1 e na F_2 e em quais proporções se aves com crista única forem cruzados com aves de uma linhagem pura em noz?

b. Quais são os genótipos dos genitores em um cruzamento noz × rosa, a partir do qual a progênie é $\frac{3}{8}$ rosa, $\frac{3}{8}$ noz, $\frac{1}{8}$ ervilha e $\frac{1}{8}$ única?

c. Quais são os genótipos dos genitores em um cruzamento noz × rosa, a partir do qual toda a progênie é noz?

d. Quantos genótipos produzem um fenótipo noz? Escreva-os.

45. A produção do pigmento da cor dos olhos em *Drosophila* requer o alelo dominante *A*. O alelo dominante *P* de um segundo gene independente torna o pigmento roxo, mas seu alelo recessivo torna-o vermelho. Uma mosca que não produz pigmento apresenta olhos brancos. Duas linhagens puras foram cruzadas, com os resultados a seguir:

P — Fêmea com olhos vermelhos × Macho com olhos brancos

↓

F_1 — Fêmeas com olhos roxos; Machos com olhos vermelhos

$F_1 \times F_1$

↓

F_2 — Ambos machos e fêmeas: $\frac{3}{8}$ com olhos roxos; $\frac{3}{8}$ com olhos vermelhos; $\frac{2}{8}$ com olhos brancos

Explique esse modo de herança, e demonstre os genótipos dos genitores, da F_1 e da F_2.

46. Quando cães marrons puros são cruzados com determinados cães brancos puros, todos os filhotes da F_1 são brancos. A progênie F_2 de alguns cruzamentos $F_1 \times F_1$ é de 118 filhotes brancos, 32 pretos e 10 marrons. Qual é a base genética desses resultados?

47. Linhagens do tipo selvagem do fungo haploide *Neurospora* produzem seu próprio triptofano. Um alelo anormal *td* torna o fungo incapaz de produzir seu próprio triptofano. Um indivíduo de genótipo *td* cresce apenas quando seu meio fornece triptofano. O alelo *su* se distribui independentemente de *td*; seu único efeito conhecido é suprimir o fenótipo *td*. Portanto, linhagens que carreiam ambos, *td* e *su*, não necessitam de triptofano para o crescimento.

a. Se uma linhagem *td*; *su* for cruzada com uma linhagem genotipicamente do tipo selvagem, quais genótipos serão esperados na progênie e em quais proporções?

b. Qual será a proporção da progênie dependente do triptofano e independente do triptofano no cruzamento da parte *a*?

48. Camundongos dos genótipos *A/A*; *B/B*; *C/C*; *D/D*; *S/S* e *a/a*; *b/b*; *c/c*; *d/d*; *s/s* são cruzados. A progênie é entrecruzada. Quais fenótipos serão produzidos na F_2 e em quais proporções? (Os símbolos de alelos fazem

referência ao que segue: *A* = Aguti, *a* = Sólido [não aguti]; *B* = Pigmento preto, *b* = Marrom; *C* = Pigmentado, *c* = Albino; *D* = Nenhuma diluição, *d* = Diluição [cor leitosa]; *S* = Não manchado, *s* = Manchas pigmentadas em fundo branco.)

49. Considere os genótipos de duas linhagens de galinhas: a linhagem pura manchada hondurenha é *i/i*; *D/D*; *M/M*; *W/W*, e a linhagem pura leghorne é *I/I; d/d*; *m/m*; *w/w*, onde:

I = Penas brancas, *i* = Penas coloridas
D = Crista dupla, *d* = Crista simples
M = Com barbela, *m* = Sem barbela
W = Pele branca, *w* = Pele amarela

Esses quatro genes se distribuem independentemente. Iniciando com essas duas linhagens puras, qual é o modo mais rápido e mais conveniente de gerar uma linhagem pura que apresente penas coloridas, apresente uma crista simples, seja sem barbela e apresente pele amarela? Assegure-se de demonstrar:

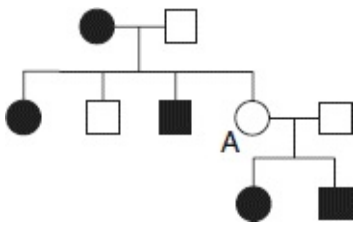
a. O heredograma do cruzamento.

b. O genótipo de cada animal representado.

c. Quantos ovos colocar para chocar em cada cruzamento e o motivo dessa quantidade.

d. Por que o seu esquema é o mais rápido e o mais conveniente.

50. O heredograma a seguir é em relação a um fenótipo dominante regulado por um alelo autossômico. O que esse heredograma sugere a respeito do fenótipo, e o que você pode deduzir a respeito do genótipo do indivíduo A?



51. A cor das pétalas em dedaleiras é determinada por três genes. *M* codifica uma enzima que sintetiza a antocianina, o pigmento roxo observado nessas pétalas; *m/m* não produz pigmento, resultando no fenótipo albino com manchas amareladas. *D* é um intensificador da antocianina, resultando em um pigmento mais escuro; *d/dia* não intensifica. No terceiro *locus*, *w/w* possibilita a deposição do pigmento nas pétalas, mas *W* evita a deposição do pigmento, exceto nas manchas e, assim, resulta no fenótipo branco e manchado. Considere os dois cruzamentos a seguir:

Cruzamento	Genitores	Progênie
1	Roxo-escuro × Branco com manchas amareladas	$\frac{1}{2}$ roxo-escuro: $\frac{1}{2}$ roxo-claro
2	Branco com manchas amareladas × Roxo-claro	$\frac{1}{2}$ branco com manchas roxas: $\frac{1}{4}$ roxo-escuro: $\frac{1}{4}$ roxo-claro

Em cada caso, forneça os genótipos dos genitores e da progênie em relação aos três genes.

52. Em uma espécie de *Drosophila*, as asas normalmente são de formato redondo, mas você obteve duas linhagens puras, uma das quais apresenta asas ovais e a outra, asas falciformes. Cruzamentos entre linhagens puras revelam os resultados a seguir:

Genitores		F_1	
Fêmea	Macho	Fêmea	Macho
falciforme	redonda	falciforme	falciforme

redonda	falciforme	falciforme	redonda
falciforme	oval	oval	falciforme

a. Forneça uma explicação genética desses resultados, definindo todos os símbolos alélicos.

b. Se as fêmeas F_1 ovais do cruzamento 3 forem cruzadas com os machos redondos da F_1 do cruzamento 2, quais proporções fenotípicas são esperadas em relação a cada sexo na progênie?

53. Camundongos normalmente apresentam uma faixa amarela em cada pelo, mas são conhecidas variantes com duas ou três faixas. Uma fêmea de camundongo que apresenta uma faixa foi cruzada com um macho que apresenta três faixas. (Nenhum animal era de uma linhagem pura.) A progênie foi:

Fêmeas	$\frac{1}{2}$ uma faixa	Machos	$\frac{1}{2}$ uma faixa
	$\frac{1}{2}$ três faixas		$\frac{1}{2}$ duas faixas

a. Forneça uma explicação clara da herança desses fenótipos.

b. De acordo com o seu modelo, qual seria o desfecho de um cruzamento entre uma filha com três faixas e um filho com uma faixa?

54. Em visons, os tipos selvagens apresentam uma pelagem quase preta. Criadores desenvolveram muitas linhagens puras de variantes da cor para a indústria de casacos de visom. Duas das referidas linhagens puras são platina (cinza-azulada) e aleutiana (cinza-aço). Essas linhagens foram utilizadas em cruzamentos, com os resultados a seguir:

Cruzamento	Genitores	F_1	F_2
1	Selvagem ×	selvagem	18 selvagens, 5

	Platina		platinas
2	Selvagem × Aleutiana	selvagem	27 selvagens, 10 aleutianas
3	Platina × Aleutiana	selvagem	133 selvagens, 41 platinas, 46 aleutianas, 17 safiras (novas)

a. Planeje uma explicação genética desses três cruzamentos. Demonstre os genótipos em relação aos genitores, à F_1, e à F_2 nos três cruzamentos, e assegure-se de demonstrar os alelos de cada gene da hipótese que você formula em relação a cada visom.

b. Preveja as proporções fenotípicas da F_1 e da F_2 a partir do cruzamento de safira com as linhagens puras platina e aleutiana.

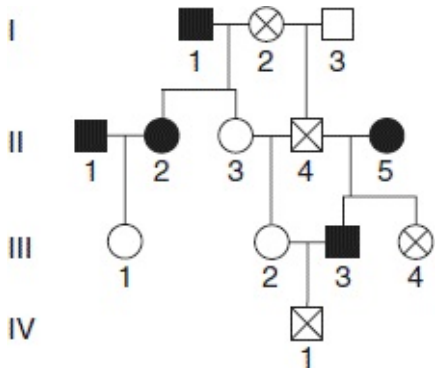
55. Na *Drosophila*, um gene autossômico determina o formato das cerdas, com *B* proporcionado cerdas retas e *b* proporcionando cerdas inclinadas. Em outro autossomo, existe um gene do qual um alelo dominante *I* inibe a formação das cerdas, de modo que a mosca é sem cerdas (*i* não apresenta efeito fenotípico conhecido).

a. Se uma mosca com cerdas retas de uma linhagem pura for cruzada com uma mosca de uma linhagem sem cerdas pura que sabidamente é um genótipo inclinado inibido, quais serão os genótipos e os fenótipos da F_1 e da F_2?

b. Qual cruzamento forneceria a proporção de 4 sem cerdas:3 retas:1 inclinada?

56. O heredograma a seguir diz respeito aos fenótipos dos olhos em besouros

Tribolium. Os símbolos sólidos representam olhos pretos, os símbolos abertos representam olhos marrons e os símbolos cruzados (X) representam o fenótipo “sem olhos”, no qual os olhos estão totalmente ausentes.



a. A partir desses dados, deduza o modo de herança desses três fenótipos.
b. Com a utilização de símbolos gênicos definidos, demonstre o genótipo do besouro II-3.

57. Uma planta que se acredita ser heterozigota em relação a um par de alelos *B/b* (no qual *B* codifica amarelo e *b* codifica bronze) foi autofecundada e, na progênie, havia 280 plantas amarelas e 120 bronze. Esses resultados amparam a hipótese de que a planta é *B/b*?

58. Uma planta que se acredita ser heterozigota em relação a dois genes de distribuição independente (*P/p*; *Q/q*) foi autofecundada, e a progênie foi:

88 *P/—*; *Q/—*	25 *p/p*; *Q/—*
32 *P/—*; *q/q*	14 *p/p*; *q/q*

Esses resultados amparam a hipótese de que a planta original era *P/p*; *Q/q*?

59. Uma planta de fenótipo 1 foi autofecundada e, na progênie, havia 100 plantas de fenótipo 1 e 60 plantas de um fenótipo alternativo 2. Esses números são compatíveis com as proporções esperadas de 9:7, 13:3 e 3:1? Forme uma hipótese genética com base nos seus cálculos.

60. Quatro linhagens mutantes homozigotas recessivas de *Drosophila melanogaster* (rotuladas 1 a 4) demonstram coordenação anormal das

patas, o que faz com que elas andem de modo altamente errático. Essas linhagens foram entrecruzadas; os fenótipos das moscas da F_1 estão demonstrados na tabela a seguir, na qual "+" representa a deambulação do tipo selvagem e "—" representa a deambulação anormal:

	1	2	3	4
1	—	+	+	+
2	+	—	—	+
3	+	—	—	+
4	+	+	+	—

a. Qual tipo de teste essa análise representa?

b. Quantos genes diferentes sofreram mutação na criação dessas quatro linhagens?

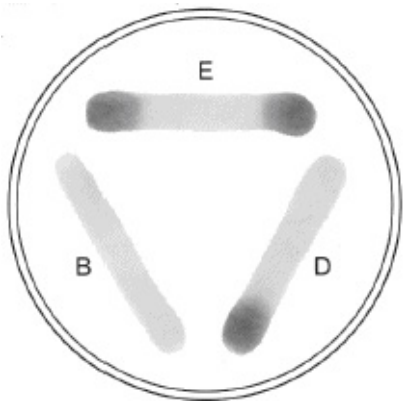
c. Invente símbolos para os tipos selvagem e mutante, e escreva os genótipos completos em relação a todas as quatro linhagens e em relação às moscas da F_1.

d. Esses dados nos informam quais genes estão ligados? Caso negativo, como a ligação pode ser testada?

e. Esses dados nos informam o número total de genes que participam na coordenação das patas nesse animal?

61. Três mutantes de uma levedura haploide que necessitam de triptofano isolados de modo independente são denominadas *trpB*, *trpD* e *trpE*. Suspensões celulares de cada um são espalhadas em uma placa de meio nutricional suplementado com triptofano suficiente apenas para possibilitar o crescimento fraco de uma linhagem *trp*. As faixas estão dispostas em um padrão triangular, de modo que elas não se toquem. É observado crescimento exuberante em ambas as extremidades da faixa *trpE* e em uma

extremidade da faixa *trpD* (ver figura adiante).



a. Você acredita que tenha havido complementação?

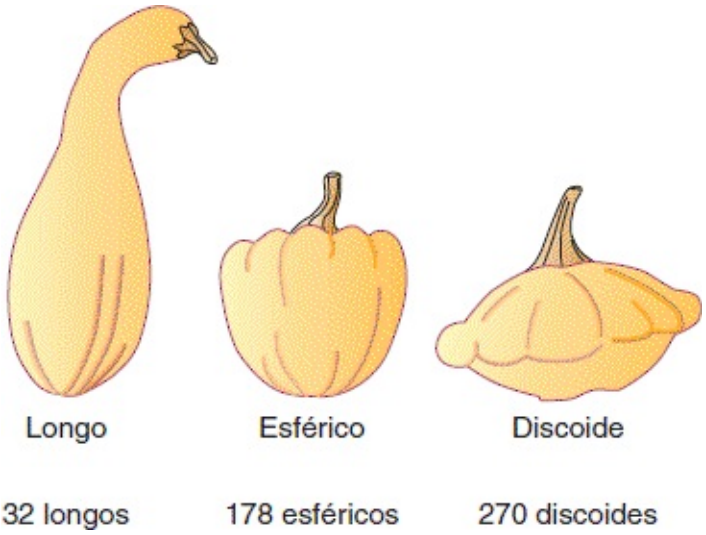
b. Explique brevemente o padrão do crescimento exuberante.

c. Desenhe as etapas enzimáticas que são defeituosas nos mutantes *trpB, trpD* e *trpE* na ordem na via de síntese do triptofano.

d. Por que foi necessário adicionar uma pequena quantidade de triptofano ao meio para demonstrar esse padrão de crescimento?

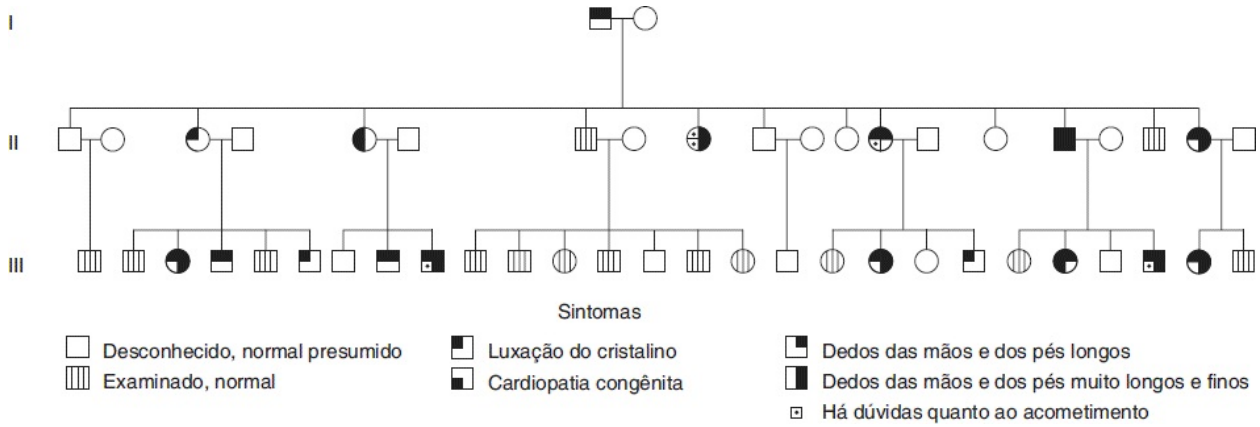
PROBLEMAS DESAFIADORES

62. Uma linhagem pura de abóbora, que produz frutos com formato de disco (ver a ilustração a seguir), foi cruzada com uma linhagem pura que apresenta frutos longos. A F_1 apresentou frutos em disco, mas a F_2 demonstrou um novo fenótipo, esférico, e foi composta pelas proporções a seguir:



Proponha uma explicação para esses resultados e demonstre os genótipos das gerações P, F_1 e F_2.

63. A síndrome de Marfan é um distúrbio do tecido conjuntivo fibroso, caracterizada por muitos sintomas, incluindo dedos longos e finos; defeitos oculares; cardiopatia e membros longos. (Flo Hyman, a estrela do voleibol americano, sofria de síndrome de Marfan. Ela morreu em virtude de uma ruptura da aorta.)



a. Utilize o heredograma anterior para propor um modo de herança em relação à síndrome de Marfan.

b. Qual fenômeno genético é demonstrado por esse heredograma?

c. Especule um motivo para esse fenômeno.

(Dados de J. V. Neel e W. J. Schull, *Human Heredity*. University of

Chicago Press, 1954.)

64. No milho, três alelos dominantes, denominados *A, C* e *R*, são necessários para produzir sementes coloridas. O genótipo *A/—*; *C/—*; *R/—* é colorido; todos os outros são incolores. Uma planta colorida é cruzada com três plantas testadoras de genótipo conhecido. Com a testadora *a/a*; *c/c*; *R/R*, a planta colorida produz 50% de sementes coloridas com *a/a*; *C/C*; *r/r*, ela produz 25% de coloridas; e com *A/A*; *c/c*; *r/r*, ela produz 50% de coloridas. Qual é o genótipo da planta colorida?

65. A produção do pigmento na camada exterior das sementes do milho requer que cada um dos três genes de distribuição independente *A, C* e *R* seja representado por no mínimo um alelo dominante, conforme especificado no Problema 64. O alelo dominante *Pr* de um quarto gene de distribuição independente é necessário para converter o precursor bioquímico em um pigmento roxo, e seu alelo recessivo *pr* torna o pigmento vermelho. As plantas que não produzem pigmento apresentam sementes amarelas. Considere um cruzamento de uma linhagem de genótipo *A/A*; *C/C*; *R/R*; *pr/pr* com uma linhagem de genótipo *a/a*; *c/c*; *r/r*; *Pr/Pr*.

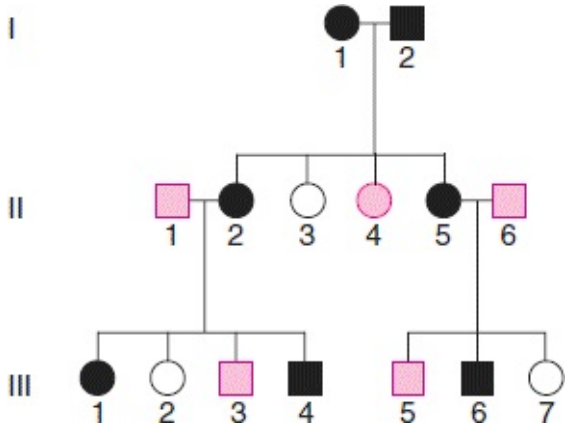
a. Quais são os fenótipos dos genitores?

b. Qual será o fenótipo da F_1?

c. Quais fenótipos, e em quais proporções, aparecerão na progênie de uma F_1 autofecundada?

d. Quais proporções da progênie você prevê a partir do cruzamento-teste de uma F_1?

66. O alelo *B* proporciona aos camundongos uma pelagem preta e *b* proporciona uma pelagem marrom. O genótipo *e/e* de outro gene de distribuição independente evita a expressão de *B* e *b*, tornando a cor da pelagem bege, enquanto *E/—* possibilita a expressão de *B* e *b*. Ambos os genes são autossômicos. No heredograma a seguir, os símbolos pretos indicam pelagem preta, os símbolos rosa indicam marrom, e os símbolos brancos indicam bege.



a. Qual é a denominação dada ao tipo de interação gênica nesse exemplo?
b. Quais são os genótipos dos camundongos no heredograma? (Se existem possibilidades alternativas, declare-as.)

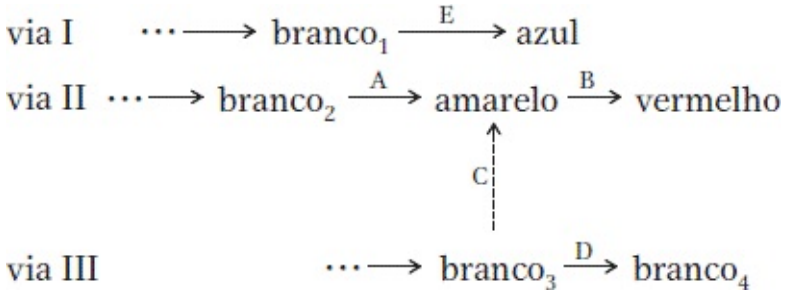
67. Um pesquisador cruza duas linhagens com flores brancas de plantas *Antirrhinum* e obtém os resultados a seguir:

Linhagem pura 1 × Linhagem pura 2
↓
F_1 todas brancas
$F_1 \times F_1$
↓
F_2 131 brancas
29 vermelhas

a. Deduza a herança desses fenótipos; utilize símbolos gênicos claramente definidos. Forneça os genótipos dos genitores, da F_1 e da F_2.
b. Preveja o resultado dos cruzamentos da F_1 com cada linhagem parental.

68. Presuma que dois pigmentos, vermelho e azul, são misturados para fornecer a cor roxa normal das pétalas de petúnia. Vias bioquímicas separadas sintetizam os dois pigmentos, conforme demonstrado nas duas fileiras superiores do diagrama a seguir. “Branco” se refere aos compostos que não são pigmentos. (A ausência total de pigmento resulta em uma pétala branca.) O pigmento vermelho é formado a partir de um intermediário

amarelo que normalmente está a uma concentração muito baixa para colorir as pétalas.



Uma terceira via, cujos compostos não contribuem com pigmento para as pétalas, normalmente não afeta as vias azul e vermelha, mas, se um de seus intermediários ($branco_3$) aumentar em concentração, pode ser convertido no intermediário amarelo da via vermelha.

No diagrama, as letras A a E representam enzimas; seus genes correspondentes, todos os quais não estão ligados, podem ser simbolizados pelas mesmas letras.

Presuma que os alelos do tipo selvagem sejam dominantes e codifiquem a função enzimática e que os alelos recessivos resultem em uma ausência de função enzimática. Deduza quais combinações de genótipos parentais verdadeiros podem ser cruzadas para produzir a progênie F_2 nas proporções a seguir:

a. 9 roxas:3 verdes:4 azuis

b. 9 roxas:3 vermelhas:3 azuis:1 branca

c. 13 roxas:3 azuis

d. 9 roxas:3 vermelhas:3 verdes:1 amarela

(Nota: azul misturado com amarelo produz verde; presuma que nenhuma mutação seja letal.)

69. As flores de capuchinha (*Tropaeolum majus*) podem ser únicas (S), duplas (D) ou superduplas (Sd). As superduplas são estéreis do sexo feminino; elas tiveram origem a partir de uma variedade com flores duplas. Cruzamentos entre variedades forneceram a progênie listada na tabela a

seguir, na qual *pura* significa “homozigoto”.

Cruzamento	Genitores	Progênie
1	S pura × D pura	Todas S
2	F_1 do cruzamento 1 × F_1 do cruzamento 1	78 S: 27 D
3	D pura × Sd	112 Sd: 108 D
4	S pura × Sd	8 Sd: 7 S
5	D pura × progênie Sd do cruzamento 4	18 Sd: 19 S
6	D pura × progênie S do cruzamento 4	14 D: 16 S

Com a utilização dos seus próprios símbolos genéticos, proponha uma explicação para esses resultados, demonstrando:

a. Todos os genótipos em cada uma das seis fileiras.

b. A origem proposta da superdupla.

70. Em uma determinada espécie de moscas, a cor dos olhos normal é vermelha (R). Quatro fenótipos anormais em relação à cor dos olhos foram observados: dois eram amarelos (Y1 e Y2), um era marrom (B) e um laranja (O). Foi estabelecida uma linhagem pura em relação a cada fenótipo, e todas as combinações possíveis das linhagens puras foram cruzadas. Moscas de cada F_1 foram entrecruzadas para produzir uma F_2. As moscas da F_1 e da F_2 estão demonstradas no quadrado a seguir; as linhagens puras são fornecidas na parte superior e do lado esquerdo.

		Y1	Y2	B	O
Y1	F_1	todas Y	todas R	todas R	todas R
	F_2	todas Y	9 R	9 R	9 R
			7 Y	4 Y	4 O
				3 B	3 Y
Y2	F_1		todas Y	todas R	todas R
	F_2		todas Y	9 R	9 R
				4 Y	4 Y
				3 B	3 O
B	F_1			todas B	todas R
	F_2			todas B	9 R
					4 O
					3 B
O	F_1				todas O
	F_2				todas O

a. Defina seus próprios símbolos e liste os genótipos de todas as quatro linhagens puras.

b. Demonstre como os fenótipos da F_1 e as proporções da F_2 são

produzidos.

c. Demonstre uma via bioquímica que explique os resultados genéticos, indicando qual gene controla qual enzima.

71. No trigo comum, *Triticum aestivum*, a cor da espiga é determinada por genes multiplamente duplicados, cada um com um alelo *R* e um *r*. Qualquer número de alelos *R* proporcionará vermelho, e uma ausência completa de alelos *R* proporcionará o fenótipo branco. Em um cruzamento entre uma linhagem pura vermelha e uma linhagem pura branca, a F_2 foi de $\frac{63}{64}$ vermelhos e $\frac{1}{64}$ branco.

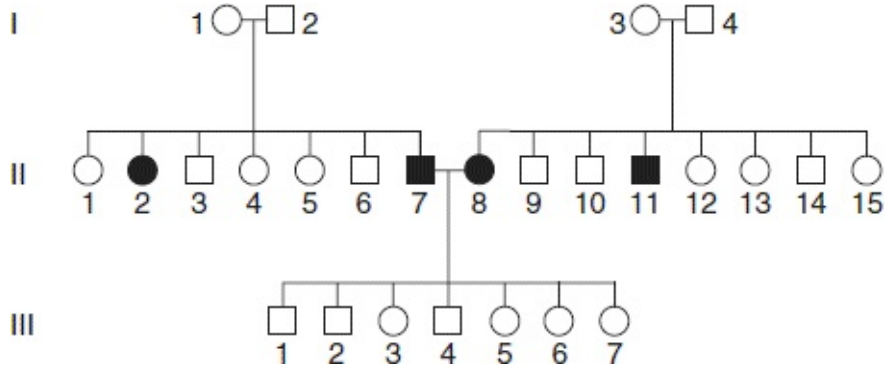
a. Quantos genes R estão segregando nesse sistema?

b. Demonstre os genótipos dos genitores, da F_1 e da F_2.

c. Plantas F_2 diferentes são retrocruzadas com o progenitor branco. Forneça exemplos de genótipos que forneceriam as proporções de progênie a seguir nos referidos retrocruzamentos: (1) 1 vermelha:1 branca, (2) 3 vermelhas:1 branca, (3) 7 vermelhas:1 branca.

d. Qual é a fórmula que em geral relaciona o número de genes em segregação com a proporção de indivíduos vermelhos na F_2 nos referidos sistemas?

72. O heredograma a seguir demonstra a herança do surdo-mutismo.

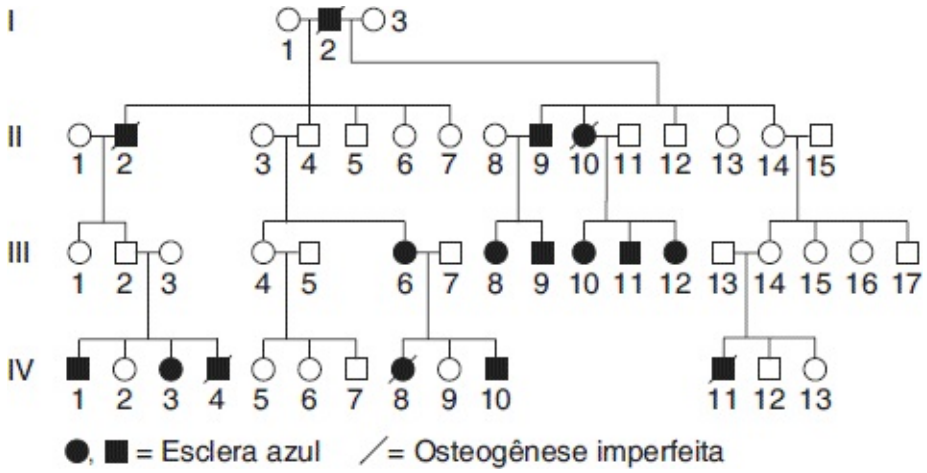


a. Forneça uma explicação para a herança dessa condição rara nas duas famílias nas gerações I e II, demonstrando os genótipos de tantas pessoas quanto possível; utilize símbolos de sua própria escolha.

b. Forneça uma explicação para a produção de apenas pessoas normais na

geração III, assegurando-se de que a sua explicação seja compatível com a resposta da parte *a*.

73. O heredograma a seguir é em relação à esclera azul e osteogênese imperfeita.



a. Essas duas anormalidades são causadas pelo mesmo gene ou por genes separados? Declare claramente os seus motivos.

b. O gene (ou os genes) é (são) autossômico(s) ou ligado(s) ao sexo?

c. O heredograma demonstra alguma evidência de penetrância incompleta ou expressividade? Caso afirmativo, realize os melhores cálculos que você conseguir dessas medidas.

74. As operárias da linhagem de abelhas melíferas conhecida como *Marrom* (absolutamente não relacionada com a cor) demonstram o que é denominado "comportamento higiênico"; ou seja, elas abrem as tampas de compartimentos da colmeia que contêm pupas mortas e em seguida removem as pupas mortas. Esse comportamento evita a difusão de bactérias infecciosas pela colônia. As operárias da linhagem *Van Scoy*, entretanto, não realizam essas tarefas e, portanto, diz-se que essa linhagem é "não higiênica". Quando uma rainha da linhagem *Marrom* foi cruzada com zangões *Van Scoy*, toda a F_1 foi não higiênica. Quando zangões dessa F_1 inseminaram uma rainha da linhagem *Marrom*, os comportamentos da progênie foram como segue:

$\frac{1}{4}$ higiênico

$\frac{1}{4}$ de abertura das tampas, mas sem a remoção das pupas

$\frac{1}{2}$ não higiênico

Entretanto, quando a tampa do compartimento de pupas mortas foi aberta pelo apicultor e as abelhas melíferas não higiênicas foram adicionalmente examinadas, observou-se que aproximadamente metade das abelhas removia as pupas mortas, mas a outra metade não removia.

a. Proponha uma hipótese genética para explicar esses padrões comportamentais.

b. Discuta os dados em relação a epistasia, dominância e interação ambiental.

(Nota: as operárias são estéreis, e todas as abelhas de uma linhagem carreiam os mesmos alelos.)

75. A cor normal das bocas-de-leão é vermelha. Foram encontradas algumas linhagens puras que demonstram variações da cor das flores. Quando essas linhagens puras foram cruzadas, elas forneceram os resultados a seguir (ver tabela):

Cruzamento	Genitores	F_1	F_2
1	Laranja × Amarela	laranja	3 laranja: 1 amarela
2	Vermelha × Laranja	vermelha	3 vermelhas: 1 laranja
3	Vermelha × Amarela	vermelha	3 vermelhas: 1 amarela
4	Vermelha × Branca	vermelha	3 vermelhas: 1 branca

5	Amarela × Branca	vermelha	9 vermelhas: 3 amarelas: 4 brancas
6	Laranja × Branca	vermelha	9 vermelhas: 3 laranja: 4 brancas
7	Vermelha × Branca	vermelha	9 vermelhas: 3 laranja: 4 brancas

a. Explique a herança dessas cores.

b. Escreva os genótipos dos genitores, da F_1 e da F_2.

76. Considere os indivíduos da F_1 a seguir em diferentes espécies e as proporções da F_2 produzida por autofecundação:

F_1	Proporção fenotípica na F_2		
1 creme	$\frac{12}{16}$ creme	$\frac{3}{16}$ preto	$\frac{1}{16}$ cinza
2 laranja	$\frac{9}{16}$ laranja	$\frac{7}{16}$ amarelo	
3 preta	$\frac{13}{16}$ preto	$\frac{3}{16}$ branco	
4 vermelho sólido	$\frac{9}{16}$ vermelho sólido	$\frac{3}{16}$ vermelho manchado	$\frac{4}{16}$ com pequenos pontos vermelhos

Se cada F_1 fosse submetida ao cruzamento-teste, quais proporções fenotípicas resultariam na progênie do cruzamento-teste?

77. Para compreender a base genética da locomoção no nematódeo diploide *Caenorhabditis elegans*, foram obtidas mutações recessivas, todas fazendo com que o parasita "oscile" de modo ineficaz, em vez de se movimentar com seu movimento habitual de deslizamento suave. Essas mutações presumivelmente afetam o sistema nervoso ou muscular. Doze mutantes

homozigotos foram entrecruzados, e os híbridos da F_1 foram examinados para verificar se eles oscilavam. Os resultados foram como segue, em que um sinal de mais significa que o híbrido da F_1 era do tipo selvagem (deslizante) e "w" significa que o híbrido oscilava.

	1	2	3	4	5	6	7	8	9	10	11	12
1	w	+	+	+	w	+	+	+	+	+	+	+
2		w	+	+	+	w	+	w	+	w	+	+
3			w	w	+	+	+	+	+	+	+	+
4				w	+	+	+	+	+	+	+	+
5					w	+	+	+	+	+	+	+
6						w	+	w	+	w	+	+
7							w	+	+	+	w	w
8								w	+	w	+	+
9									w	+	+	+
10										w	+	+
11											w	w
12												w

a. Explique o que esse experimento foi desenhado para testar.

b. Utilize essa justificativa para atribuir os genótipos para todos os 12

mutantes.

c. Explique por que o fenótipo dos híbridos da F_1 entre os mutantes 1 e 2 diferiu daquele dos híbridos entre os mutantes 1 e 5.

78. Um geneticista que trabalha com um fungo haploide realiza um cruzamento entre dois mutantes de crescimento lento, denominados *musgoso* e *aranha* (fazendo referência ao aspecto anormal das colônias). As tétrades do cruzamento são de três tipos (A, B, C), mas duas delas contêm esporos que não germinam.

Esporo	A	B	C
1	tipo selvagem	tipo selvagem	aranha
2	tipo selvagem	aranha	aranha
3	nenhuma germinação	musgoso	musgoso
4	nenhuma germinação	nenhuma germinação	musgoso

Planeje um modelo para explicar esses resultados genéticos e proponha uma base molecular para o seu modelo.

79. No nematódeo *C. elegans*, alguns parasitas apresentam cutículas bolhosas em virtude de uma mutação recessiva em um dos genes *bli*. Pessoas que estudavam uma mutação supressora que suprimia as mutações *bli-3* desejavam saber se ela também suprime as mutações em *bli-4*. Elas possuíam uma linhagem homozigota em relação a essa mutação supressora recessiva, e seu fenótipo era do tipo selvagem.

a. Como elas determinariam se essa mutação supressora recessiva suprime as mutações em *bli-4*? Em outras palavras, qual é o genótipo dos parasitas

necessário para responder à pergunta?

b. Qual(is) cruzamento(s) elas realizariam para produzir esses parasitas?

c. Quais resultados elas esperariam na F_2 se

(1) ela atuasse como uma supressora de *bli-4*?

(2) ela não atuasse como uma supressora de *bli-4*?

CAPÍTULO

7

DNA | Estrutura e Replicação

Modelo em computador do DNA. (Kenneth Eward/Science Source/Getty Images.)

TÓPICOS

7.1 DNA | O material genético
7.2 Estrutura do DNA
7.3 Replicação semiconservativa
7.4 Visão geral da replicação do DNA
7.5 Replissomo | Uma máquina de replicação notável
7.6 Replicação em organismos eucariotos
7.7 Telômeros e telomerase | Término da replicação

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Avaliar os tipos de evidências (históricas e modernas) que podem ser utilizadas para demonstrar que o DNA é o material genético
- Avaliar os dados utilizados para construir o modelo de dupla-hélice do DNA
- Explicar por que a estrutura helicoidal dupla sugere um mecanismo específico de replicação do DNA
- Ilustrar as características da replicação do DNA que contribuem para a sua velocidade e a sua precisão
- Explicar por que as extremidades dos cromossomos necessitam de replicação especial
- Prever as possíveis consequências para a saúde humana se a replicação da extremidade for defeituosa.

James Watson (um geneticista microbiano americano) e Francis Crick (um físico inglês) solucionaram a estrutura do DNA em 1953. Seu modelo da estrutura do DNA foi revolucionário. Eles propuseram uma definição para o gene em termos químicos e, ao fazer isso, propiciaram a compreensão da ação gênica e da hereditariedade no nível molecular. Uma medida da importância da sua descoberta é que a estrutura em dupla-hélice se tornou um ícone cultural, que

é observado com cada vez mais frequência em pinturas, em esculturas e até mesmo em *playgrounds* (Figura 7.1).

A história tem início na primeira metade do século 20, quando os resultados de diversos experimentos levaram os cientistas a concluírem que o DNA é o material genético, não outra molécula biológica, tal como um carboidrato, uma proteína ou um lipídio. O DNA é uma molécula simples, composta por apenas quatro diferentes elementos estruturais (as quatro bases de nucleotídios). Portanto, era necessário compreender como essa molécula tão simples podia ser responsável pela incrível diversidade de organismos sobre a Terra.

O modelo da dupla-hélice proposto por Watson e Crick foi construído com base nos resultados de cientistas antes deles. Eles se basearam nas descobertas anteriores da composição química do DNA e das proporções de suas bases. Além disso, imagens de difração de raios X do DNA revelaram ao olho treinado que o DNA é uma hélice de dimensões precisas. Watson e Crick concluíram que o DNA é uma dupla-hélice composta por dois filamentos de bases de nucleotídios ligados, que se enrolam um ao redor do outro.

A estrutura do material hereditário proposta imediatamente sugeriu como ela poderia atuar como um molde e como poderia ser transmitida entre as gerações. Primeiramente, a informação para a formação de um organismo está codificada na sequência de bases nucleotídicas que compõem os dois filamentos da hélice de DNA. Em segundo lugar, em virtude das regras da complementaridade de bases descobertas por Watson e Crick, a sequência de um filamento dita a sequência do outro filamento. Assim, as informações genéticas na sequência do DNA podem ser transmitidas de uma geração para a próxima quando cada um dos filamentos de DNA em separado atua como molde para a produção de novas cópias da molécula.

Escultura do DNA

FIGURA 7.1 (Neil Grant/Alamy.)

Neste capítulo, enfocamos o DNA, sua estrutura e a produção de cópias do DNA em um processo denominado replicação. Precisamente como o DNA é replicado ainda é uma área ativa de pesquisas, mais de 50 anos após a descoberta da dupla-hélice. Nossa atual compreensão sobre o mecanismo da replicação confere um papel central para uma estrutura proteica, denominada replissomo. Esse complexo de proteínas coordena as diversas reações que são necessárias para a replicação rápida e precisa do DNA.

7.1 DNA | O material genético

Antes de verificarmos como Watson e Crick solucionaram a estrutura do DNA, revisaremos o que era conhecido a respeito dos genes e do DNA na ocasião em

que eles iniciaram a sua colaboração histórica:

1. Os genes — os "fatores" hereditários descritos por Mendel — sabidamente estavam associados a traços específicos, mas a sua natureza física não era compreendida. De modo semelhante, as mutações sabidamente alteravam a função dos genes, mas a natureza química precisa de uma mutação não era compreendida.
2. A hipótese um gene—um polipeptídio (descrita no Capítulo 6) postulava que os genes determinam a estrutura das proteínas e de outros polipeptídios.
3. Sabidamente os genes eram carreados nos cromossomos.
4. Observou-se que os cromossomos são compostos por DNA e proteínas.
5. Os resultados de uma série de experimentos com início na década de 1920 revelaram que o DNA é o material genético. Esses experimentos, descritos em seguida, demonstraram que as células bacterianas que expressam um fenótipo podem ser transformadas em células que expressam um fenótipo diferente e que o agente transformante é o DNA.

Descoberta da transformação

Frederick Griffith fez uma observação enigmática durante os experimentos realizados em 1928 com a bactéria *Streptococcus pneumoniae*. Essa bactéria, que causa pneumonia em seres humanos, normalmente é letal em camundongos. Entretanto, algumas linhagens dessa espécie bacteriana evoluíram para serem menos virulentas (menos capazes de causar doença ou morte). Os experimentos de Griffith estão resumidos na Figura 7.2. Nesses experimentos, Griffith utilizou duas linhagens que são distinguíveis pelo aspecto de suas colônias quando cultivadas em culturas de laboratório. Uma linhagem era um tipo virulento normal, mortal para a maior parte dos animais de laboratório. As células dessa linhagem estão envoltas por uma cápsula de polissacarídios, o que proporciona às colônias um aspecto liso; portanto, essa linhagem é identificada como *S*. A outra linhagem de Griffith era um tipo não virulento mutante, que cresce em camundongos, mas que não é letal. Nessa linhagem, revestimento de polissacarídios não existe, proporcionando às colônias um aspecto rugoso; essa linhagem é denominada *R*.

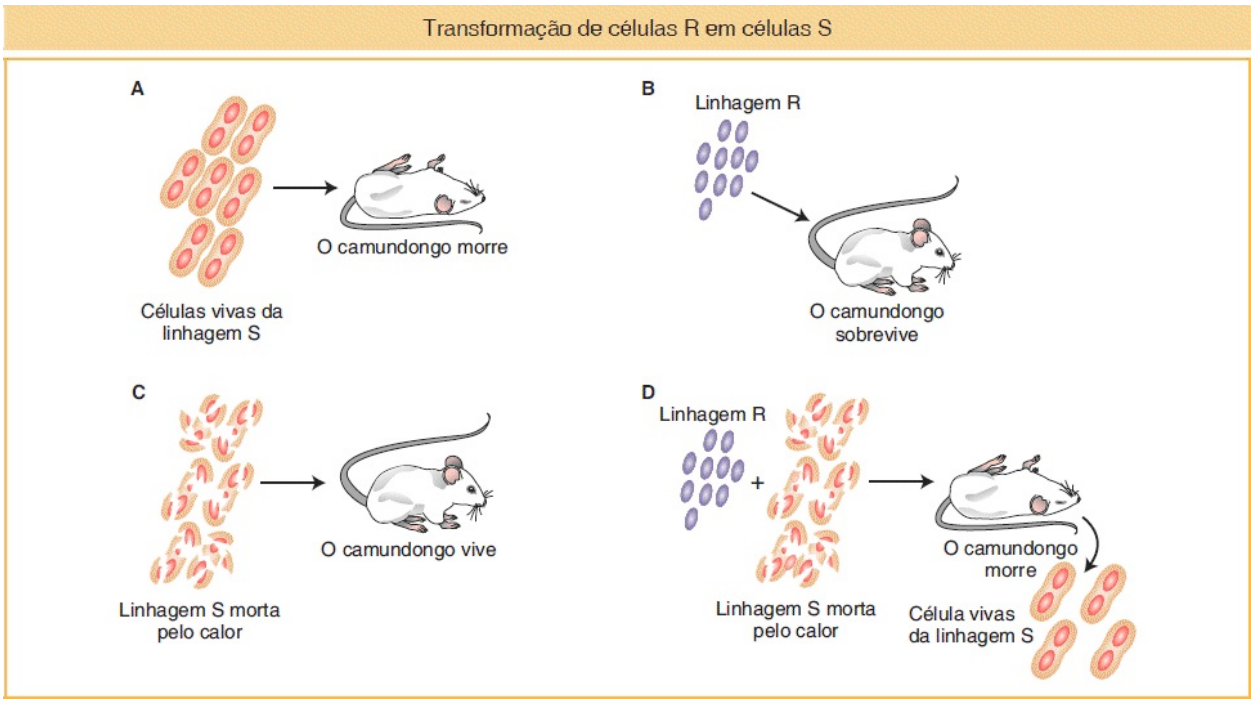


FIGURA 7.2 A presença de células S mortas pelo calor transforma as células R vivas em células S vivas. **A.** O camundongo morre após a injeção da linhagem S virulenta. **B.** O camundongo sobrevive após a injeção da linhagem R. **C.** O camundongo sobrevive após a injeção da linhagem S morta pelo calor. **D.** O camundongo morre após a injeção de uma mistura da linhagem S morta pelo calor e da linhagem R viva. Células S vivas foram isoladas do camundongo morto, indicando que a linhagem S morta pelo calor de algum modo transforma a linhagem R na linhagem S virulenta.

Griffith matou algumas células virulentas por meio de fervura. Em seguida, ele injetou as células mortas pelo calor em camundongos. Os camundongos sobreviveram, demonstrando que os restos das células não causam a morte. Entretanto, camundongos injetados com uma mistura de células virulentas mortas pelo calor e células não virulentas vivas morreram. Além disso, células vivas puderam ser recuperadas dos camundongos mortos; essas células formaram colônias lisas e eram virulentas na injeção subsequente. De algum modo, os restos celulares das células S fervidas haviam convertido as células R vivas em células S vivas. O processo, já discutido no Capítulo 5, é denominado *transformação*.

A etapa seguinte foi determinar qual componente químico das células doadoras mortas havia causado essa transformação. Essa substância modificou o genótipo da linhagem receptora e, portanto, podia ser uma candidata para o material hereditário. Esse problema foi solucionado por experimentos conduzidos em

1944 por Oswald Avery e dois colegas, Colin MacLeod e Maclyn McCarty (Figura 7.3). Sua abordagem para o problema foi destruir quimicamente todas as principais categorias de substâncias químicas em um extrato de células mortas, uma por vez, e descobrir se o extrato havia perdido a capacidade de transformação. As células virulentas apresentavam um revestimento liso de polissacarídios, enquanto as células não virulentas não apresentavam; portanto, os polissacarídios eram candidatos óbvios a serem o agente transformante. Entretanto, quando os polissacarídios foram destruídos, a mistura ainda podia ser transformada. Proteínas, lipídios e ácidos ribonucleicos (RNA) demonstraram, todos, de modo semelhante, não serem o agente transformante. A mistura perdeu a sua capacidade de transformação apenas quando a mistura doadora foi tratada com a enzima desoxirribonuclease (DNase), que fragmenta o DNA. Esses resultados implicam fortemente o DNA como o material genético. Agora se sabe que os fragmentos do DNA transformante que conferem virulência entram no cromossomo bacteriano e substituem seus correspondentes que conferem não virulência.

> CONCEITO-CHAVE A demonstração de que o DNA é o princípio transformante foi a primeira demonstração de que os genes (o material hereditário) são compostos por DNA.

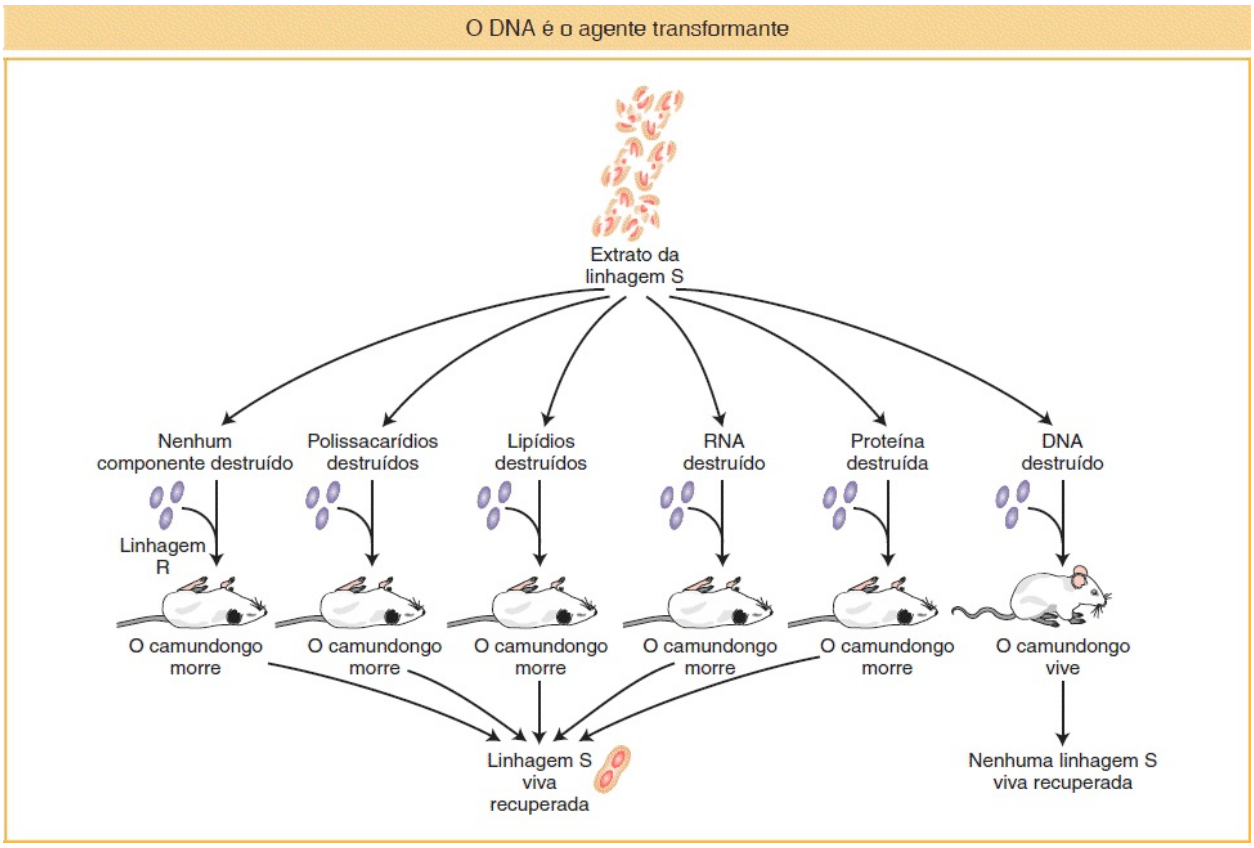


FIGURA 7.3 O DNA é o agente que transforma a linhagem R em virulenta. Se o DNA em um extrato de células da linhagem S mortas pelo calor for destruído, os camundongos sobrevivem quando injetados com uma mistura das células mortas pelo calor e as células da linhagem R não virulenta vivas.

Experimento de Hershey-Chase

Os experimentos conduzidos por Avery e seus colegas foram definitivos, mas muitos cientistas estavam relutantes em aceitar o DNA (em vez das proteínas) como o material genético. Afinal, como uma molécula de tal baixa complexidade como o DNA poderia codificar a diversidade da vida sobre este planeta? Alfred Hershey e Martha Chase forneceram evidências adicionais em 1952 em um experimento que fez uso do fago T2, um vírus que infecta bactérias. Eles ponderaram que o fago infectante obrigatoriamente injetavam na bactéria as informações específicas que ditam a reprodução de novas partículas virais. Se eles pudessem descobrir qual material o fago estava injetando na hospedeira bacteriana, eles teriam determinado o material genético dos fagos.

O fago é relativamente simples em constituição molecular. A estrutura do T2 é

similar à do T4 demonstrada nas Figuras 5.22 a 5.24. A maior parte de sua estrutura é de proteínas, com o DNA contido dentro da bainha proteica de sua "cabeça". Hershey e Chase decidiram marcar diferencialmente o DNA e a proteína por meio da utilização de radioisótopos, de modo que eles pudessem seguir os dois materiais durante a infecção. O fósforo não é observado nos aminoácidos que formam os blocos de proteínas, mas é uma parte integrante do DNA; contrariamente, o enxofre está presente nas proteínas, mas nunca no DNA. Hershey e Chase incorporaram o radioisótopo do fósforo (^{32}P) no DNA do fago e aquele do enxofre (^{35}S) nas proteínas de uma cultura de fagos em separado. Conforme demonstrado na Figura 7.4, em seguida eles infectaram duas culturas de *E. coli* com muitas partículas virais por célula: uma cultura de *E. coli* recebeu o fago marcado com ^{32}P e a outra recebeu o fago marcado com ^{35}S. Após possibilitar tempo suficiente para a ocorrência da infecção, eles fragmentaram as carcaças vazias dos fagos (denominadas *fantasmas*) das células bacterianas por meio de agitação em um liquidificador de cozinha. Eles separaram as células bacterianas dos fantasmas dos fagos em uma centrífuga e em seguida mediram a radioatividade nas duas frações. Quando os fagos marcados com ^{32}P foram utilizados para infectar *E. coli*, a maior parte da radioatividade estava dentro das células bacterianas, indicando que o DNA do fago entrava nas células. Quando os fagos marcados com ^{35}S foram utilizados, a maior parte do material radioativo estava nos fantasmas dos fagos, indicando que a proteína do fago nunca entrava na célula bacteriana. A conclusão é inevitável: o DNA é o material hereditário. As proteínas dos fagos são meros envoltórios estruturais, que são descartados após a penetração do DNA viral na célula bacteriana.

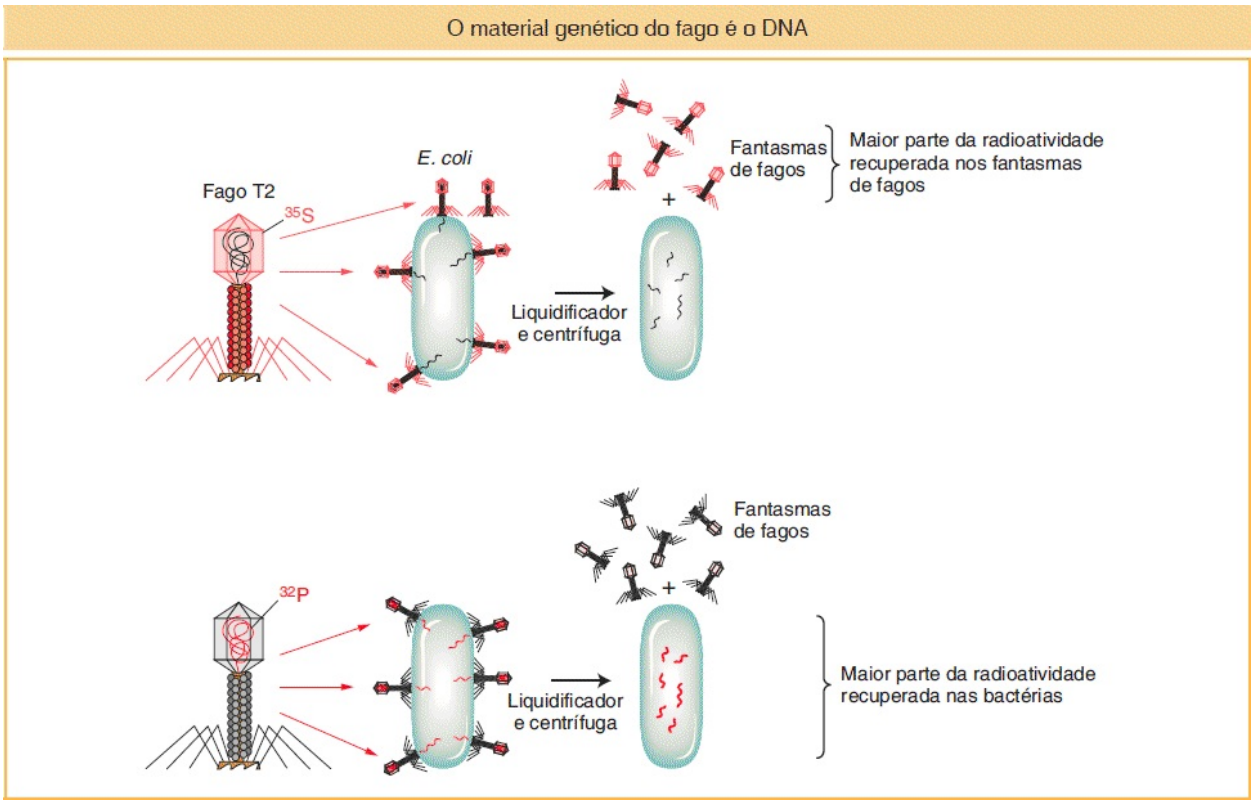


FIGURA 7.4 O experimento de Hershey-Chase demonstrou que o material genético dos fagos é o DNA, não a proteína. O experimento utiliza dois conjuntos de bacteriófagos T2. Em um conjunto, o revestimento proteico é marcado com enxofre radioativo (^{35}S), não observado no DNA. No outro conjunto, o DNA é marcado com fósforo radioativo (^{32}P), não observado em aminoácidos. Apenas ^{32}P é recuperado da *E. coli*, indicando que o DNA é o agente necessário para a produção de novos fagos.

7.2 Estrutura do DNA

Até mesmo antes de a estrutura do DNA ter sido elucidada, estudos genéticos indicavam que o material hereditário tinha necessariamente três propriedades-chave:

1. Tendo em vista que essencialmente todas as células no corpo de um organismo apresentam a mesma constituição genética, a replicação fiel do material genético a cada divisão celular é crucial. Portanto, as características estruturais do DNA *precisam assegurar a replicação fiel*. Essas características estruturais serão consideradas posteriormente neste

capítulo.

2. Tendo em vista que ele precisa codificar a constelação de proteínas expressas por um organismo, o material genético *precisa ter um conteúdo informativo*. Como as informações codificadas no DNA são decifradas para produzir proteínas será o assunto dos Capítulos 8 e 9.
3. Tendo em vista que alterações hereditárias, denominadas mutações, fornecem a matéria-prima para a seleção evolutiva, o material genético *tem de ser capaz de mudar* em ocasiões raras. Não obstante, a estrutura do DNA deve ser estável, de modo que os organismos possam "confiar" em sua informação codificada. Consideraremos os mecanismos de mutação no Capítulo 16.

Estrutura do DNA antes de Watson e Crick

Considere a descoberta da estrutura da dupla-hélice do DNA por Watson e Crick como a solução para um quebra-cabeça tridimensional complicado. Para solucionar esse quebra-cabeça, Watson e Crick utilizaram um processo denominado "construção de modelo", no qual eles reuniram os resultados de experimentos anteriores e em andamento (as peças do quebra-cabeça) para formar o quebra-cabeça tridimensional (o modelo da dupla-hélice). Para compreender como eles fizeram isso, primeiramente precisamos saber quais peças do quebra-cabeça estavam disponíveis para Watson e Crick em 1953.

Elementos estruturais do DNA. A primeira peça do quebra-cabeça era o conhecimento sobre as estruturas básicas do DNA. Como uma substância química, o DNA é consideravelmente simples. Ele contém três tipos de componentes químicos: (1) **fosfato**, (2) um açúcar denominado **desoxirribose** e (3) quatro **bases** nitrogenadas — adenina, guanina, citosina e timina. O açúcar no DNA é denominado "desoxirribose", tendo em vista que ele apresenta apenas um átomo de hidrogênio (H) no átomo de carbono 2′, contrariamente à ribose (um componente do RNA), que apresenta um grupo hidroxila (OH) naquela posição. Duas das bases, adenina e guanina, apresentam uma estrutura de anel duplo característica de um tipo de substância química denominado **purina**. As outras

duas bases, citosina e timina, apresentam uma estrutura de anel único de um tipo denominado **pirimidina**.

São atribuídos números aos átomos de carbono nas bases para facilidade de referência. Também são atribuídos números aos átomos de carbono no grupo açúcar — nesse caso, o número é seguido por apóstrofo reto (1′, 2′ e assim por diante).

Os componentes químicos do DNA estão dispostos em grupos denominados **nucleotídios**, cada um composto por um grupo fosfato, uma molécula de açúcar desoxirribose e qualquer uma das quatro bases (Figura 7.5). É uma convenção fazer referência a cada nucleotídio pela primeira letra da denominação de sua base: A, G, C ou T. O nucleotídio com a base adenina é denominado desoxiadenosina 5′-monofosfato, em que 5′ faz referência à posição do átomo de carbono no anel de açúcar ao qual o único grupo fosfato (mono) está ligado.

Regras de Chargaff da composição de bases. A segunda peça do quebra-cabeça utilizada por Watson e Crick teve origem em um trabalho realizado muitos anos antes por Erwin Chargaff. Estudando uma grande seleção de DNA de diferentes organismos (Tabela 7.1), Chargaff estabeleceu determinadas regras empíricas a respeito das quantidades de cada tipo de nucleotídio observado no DNA:

1. A quantidade total de nucleotídios pirimidínicos (T + C) sempre é igual à quantidade total de nucleotídios purínicos (A + G).
2. A quantidade de T sempre é igual à quantidade de A e a quantidade de C sempre é igual à quantidade de G. Mas a quantidade de A + T não é necessariamente igual à quantidade de G + C, conforme pode ser observado na coluna à direita da Tabela 7.1. Essa proporção varia entre os diferentes organismos, mas é virtualmente a mesma em diferentes tecidos do mesmo organismo.

Análise do DNA por difração de raios X. A terceira e mais controversa peça do quebra-cabeça teve origem nos dados de difração de raios X sobre a estrutura do DNA que foram coletados por Rosalind Franklin quando ela estava no laboratório de Maurice Wilkins (Figura 7.6). Nos referidos experimentos, foram disparados

raios X nas fibras de DNA e a dispersão dos raios a partir das fibras é observada por meio da captação dos raios em filme fotográfico, sobre o qual os raios X produzem pontos. O ângulo da dispersão representado por cada ponto no filme fornece informações a respeito da posição de um átomo ou de determinados grupos de átomos na molécula de DNA. Esse procedimento não é de fácil realização (ou explicação) e a interpretação dos padrões de pontos exige tratamento matemático complexo, que está além do escopo deste texto. Os dados disponíveis sugeriam que o DNA é longo e fino e que apresenta duas partes semelhantes, que são paralelas entre si e que estão localizadas ao longo do comprimento da molécula. Os dados dos raios X demonstraram que a molécula é helicoidal (semelhante a uma espiral). Sem o conhecimento de Rosalind Franklin, sua melhor radiografia foi mostrada a Watson e Crick por Maurice Wilkins e foi essa peça crucial do quebra-cabeça que possibilitou que eles deduzissem a estrutura tridimensional que poderia explicar os padrões de pontos dos raios X.

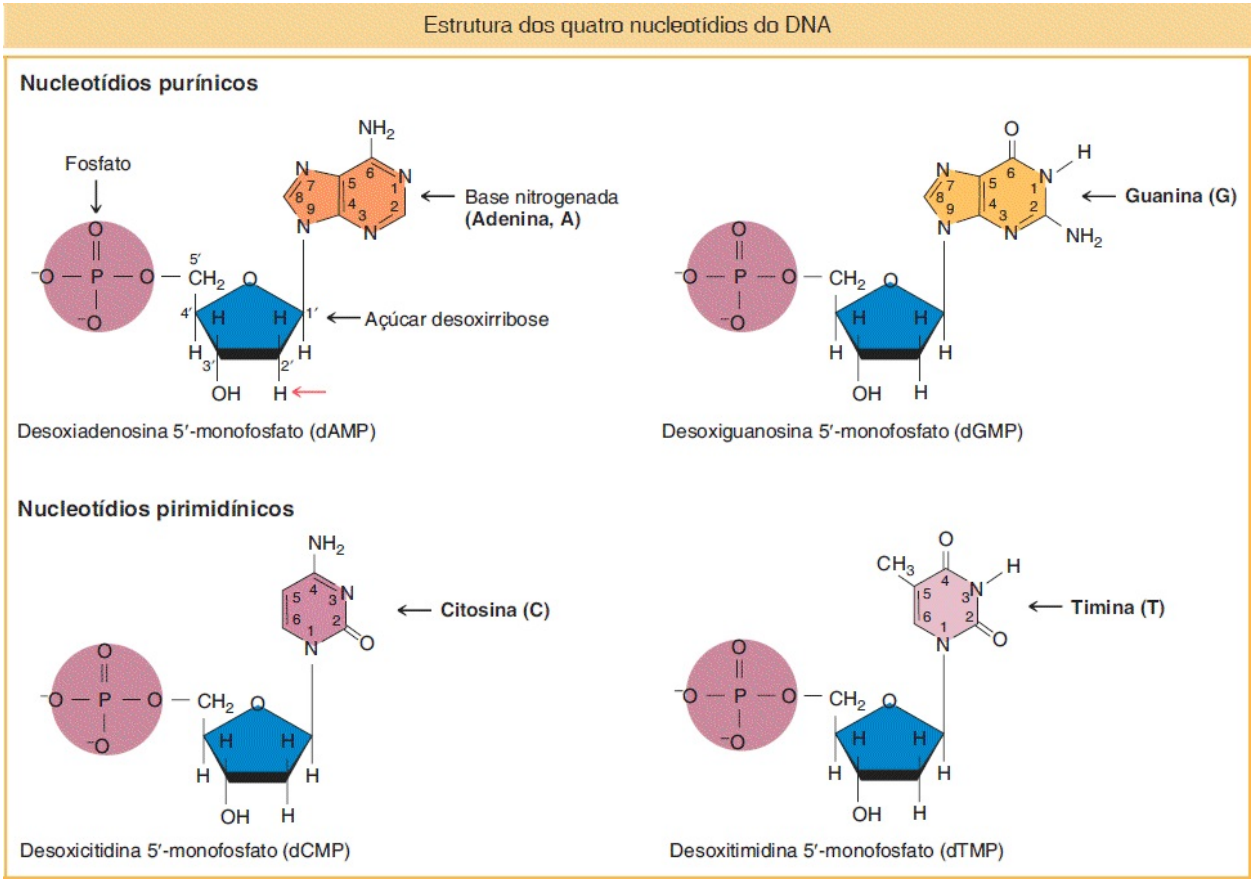


FIGURA 7.5 Estes nucleotídios, dois com bases purina e dois com bases pirimidina, são os elementos estruturais fundamentais do DNA. O açúcar é denominado *desoxirribose*, tendo em vista que é uma variação de um açúcar comum, a ribose, que contém mais um átomo de oxigênio (posição indicada pela *seta vermelha*).

Tabela 7.1 Propriedades molares das bases* em DNA de diversas fontes.

Organismo	Tecido	Adenina	Timina	Guanina	Cito
Escherichia coli (K12)	—	26,0	23,9	24,9	25,2
Diplococcus pneumoniae	—	29,8	31,6	20,5	18,0
Mycobacterium tuberculosis	—	15,1	14,6	34,9	35,4
Levedura	—	31,3	32,9	18,7	17,1
Paracentrotus lividus (ouriço-do-mar)	Espermatozoide	32,8	32,1	17,7	18,4
Arenque	Espermatozoide	27,8	27,5	22,2	22,6
Rato	Medula óssea	28,6	28,4	21,4	21,5

Ser humano	Timo	30,9	29,4	19,9	19,8
Ser humano	Fígado	30,3	30,3	19,5	19,9
Ser humano	Espermatozoide	30,7	31,2	19,3	18,8

*Definidas em moles de constituintes nitrogenados por 100 g átomos de fosfato em hidrolisado.
Fonte: Dados de E. Chargaff e J. Davidson, eds., *The Nucleic Acids*. Academic Press, 1955.

Dupla-hélice

Um artigo de 1953 de Watson e Crick no periódico *Nature* foi iniciado com duas sentenças que conduziram a uma nova era da biologia: “Desejamos sugerir uma estrutura para o sal do ácido nucleico desoxirribose (D.N.A.). Essa estrutura apresenta características novas que são de interesse biológico considerável.”[1] A estrutura do DNA havia sido tema de grande debate desde os experimentos de Avery e colaboradores em 1944. Conforme observamos, a composição geral do DNA era conhecida, mas não se sabia como as partes se encaixavam. A estrutura precisava atender as principais exigências em relação a uma molécula hereditária: a capacidade de armazenar informações, a capacidade de ser replicada e a capacidade de sofrer mutação.

A estrutura tridimensional derivada por Watson e Crick é composta por duas cadeias de nucleotídios (“filamentos”), uma ao lado da outra, torcidas no formato de uma **dupla-hélice** (Figura 7.7). Os dois filamentos de nucleotídios são mantidos unidos por ligações de hidrogênio entre as bases de cada filamento, formando uma estrutura semelhante a uma escada em espiral (Figura 7.8 A). O arcabouço de cada filamento é formado por unidades de fosfato e açúcar desoxirribose alternadas, que estão conectadas por ligações fosfodiéster (Figura 7.8 B). Podemos utilizar essas ligações para descrever como uma cadeia de nucleotídios é organizada. Conforme mencionado anteriormente, os átomos de

carbono dos grupos de açúcar são numerados 1′ a 5′. Uma ligação fosfodiéster conecta o átomo de carbono 5′ de uma desoxirribose ao átomo de carbono 3′ da desoxirribose adjacente. Portanto, diz-se que cada filamento de açúcar e fosfato apresenta uma polaridade, ou sentido, de 5′ para 3′ e a compreensão dessa polaridade é essencial para compreender como o DNA atua. Na molécula de DNA bifilamentar, os dois filamentos têm orientação oposta, ou **antiparalela** (ver Figura 7.8 B).

FIGURA 7.6 Rosalind Franklin (*esquerda*) e seu padrão de difração de raios X do DNA (*direita*). (*Esquerda*, Science Source; *direita*, Rosalind Franklin/Science Source.)

FIGURA 7.7 James Watson e Francis Crick com seu modelo do DNA. (A. Barrington Brown/Science Source.)

Cada base está ligada ao átomo de carbono 1′ de um açúcar desoxirribose no arcabouço de cada filamento e está voltada para dentro e em direção a uma base no outro filamento. Ligações de hidrogênio entre os pares de bases mantêm os dois filamentos da molécula de DNA unidos. As ligações de hidrogênio estão indicadas pelas linhas tracejadas na Figura 7.8 B.

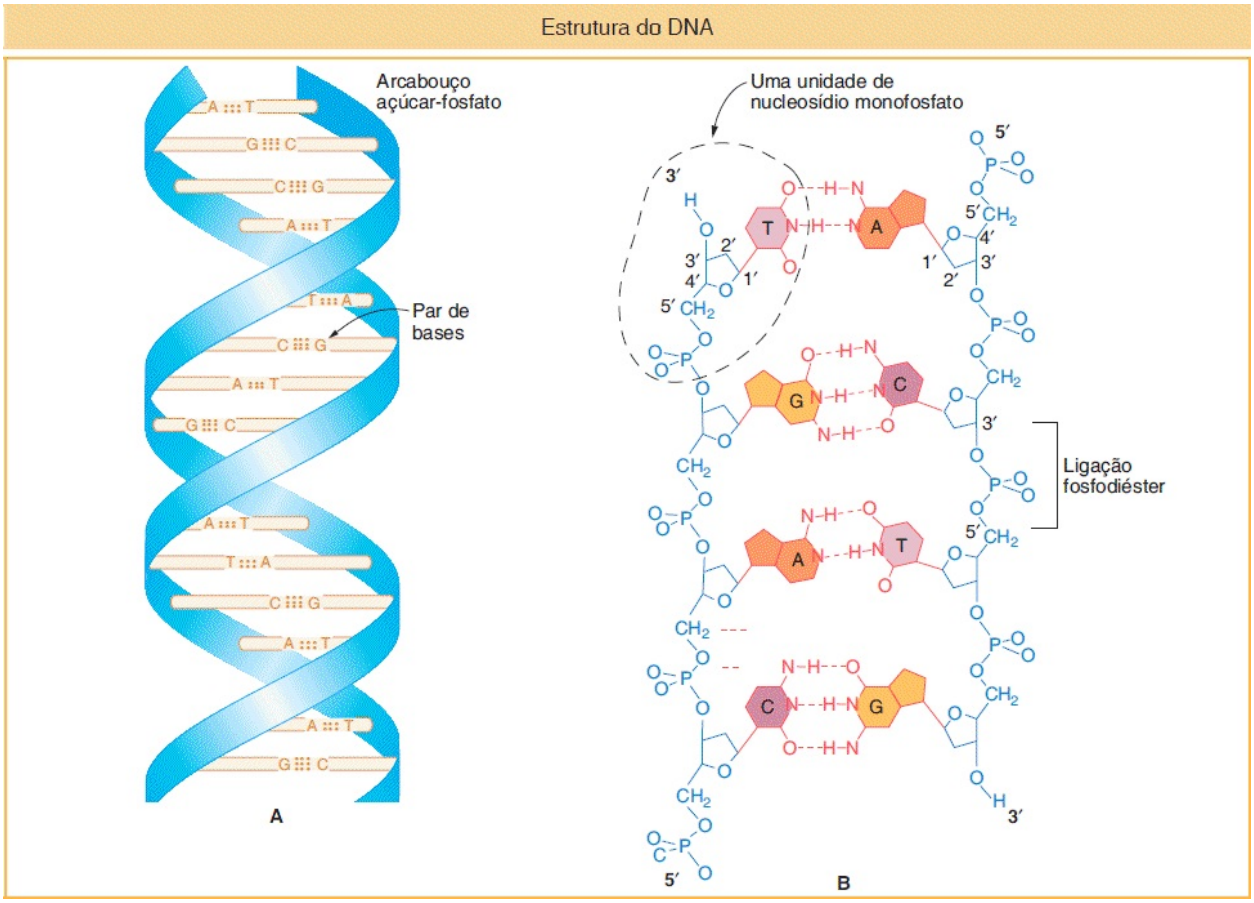


FIGURA 7.8 A. Modelo simplificado demonstrando a estrutura helicoidal do DNA. Os bastões representam os pares de bases e as fitas representam o arcabouço de açúcar e fosfato das duas cadeias antiparalelas. **B.** Diagrama químico preciso da dupla-hélice do DNA, desenrolada para demonstrar as espinhas dorsais açúcar-fosfato (*azul*) e os degraus de pares de bases (*roxo, laranja*). As espinhas dorsais correm em direções opostas; as extremidades 5′ e 3′ são denominadas em virtude da orientação dos átomos de carbono 5′ e 3′ dos anéis de açúcar. Cada par de bases apresenta uma base purina, adenina (A) ou guanina (G) e uma base pirimidina, timina (T) ou citosina (C), conectadas por ligações de hidrogênio (*linhas tracejadas vermelhas*).

Dois filamentos de nucleotídios complementares pareados de modo antiparalelo assumem automaticamente uma conformação de dupla-hélice (Figura 7.9), principalmente por meio da interação dos pares de bases. Os pares de bases, que são estruturas planares achatadas, ficam “empilhados” uns sobre os outros no centro da dupla-hélice (ver Figura 7.9 A). O empilhamento aumenta a estabilidade da molécula de DNA ao excluir as moléculas de água dos espaços entre os pares de bases. A forma mais estável que resulta do empilhamento de bases é uma dupla-hélice com dois tamanhos distintos de sulcos que ocorrem em

uma espiral: o **sulco maior** e o **sulco menor**, que podem ser observados em ambos os modelos da fita e do preenchimento dos espaços da Figura 7.9 A e 7.9 B. A maior parte das associações DNA-proteína ocorre nos sulcos maiores. Um filamento de nucleotídios único não apresenta estrutura helicoidal; o formato helicoidal do DNA depende totalmente do pareamento e do empilhamento das bases nos filamentos antiparalelos. O DNA é uma hélice com giro para a direita; em outras palavras, ele apresenta a mesma estrutura de um parafuso que seria parafusado em um local por meio de um movimento em sentido horário.

A dupla-hélice corresponde bem aos dados de difração dos raios X e às regras de Chargaff. Ao estudar os modelos da estrutura que eles produziram, Watson e Crick perceberam que o raio observado da dupla-hélice (conhecido a partir dos dados de raios X) seria explicado se uma base purina sempre pareasse (por meio de ligações de hidrogênio) com uma base pirimidina (Figura 7.10). O referido pareamento explicaria a regularidade de (A + G) = (T + C) observada por Chargaff, mas preveria quatro pareamentos possíveis: T···A, T···G, C···A e C···G. Os dados de Chargaff, entretanto, indicam que T pareia apenas com A e que C pareia apenas com G. Watson e Crick concluíram que cada par de bases é composto por uma base purina e uma base pirimidina, pareadas de acordo com a regra a seguir: G pareia com C e A pareia com T.

Observe que o par G-C apresenta três ligações de hidrogênio, enquanto o par A-T apresenta apenas duas (ver Figura 7.8 B). Preveríamos que o DNA que contém muitos pares G-C seria mais estável do que o DNA que contém muitos pares A-T. De fato, essa previsão é confirmada. O calor causa a separação dos dois filamentos da dupla-hélice do DNA (um processo denominado dissociação do DNA ou desnaturação do DNA); DNA com mais alto conteúdo de G + C necessita de temperaturas mais altas para se dissociar, em virtude da maior atração do pareamento G-C.

> CONCEITO-CHAVE O DNA é uma dupla-hélice composta por duas cadeias de nucleotídios unidas por meio do pareamento complementar de A com T e de G com C.

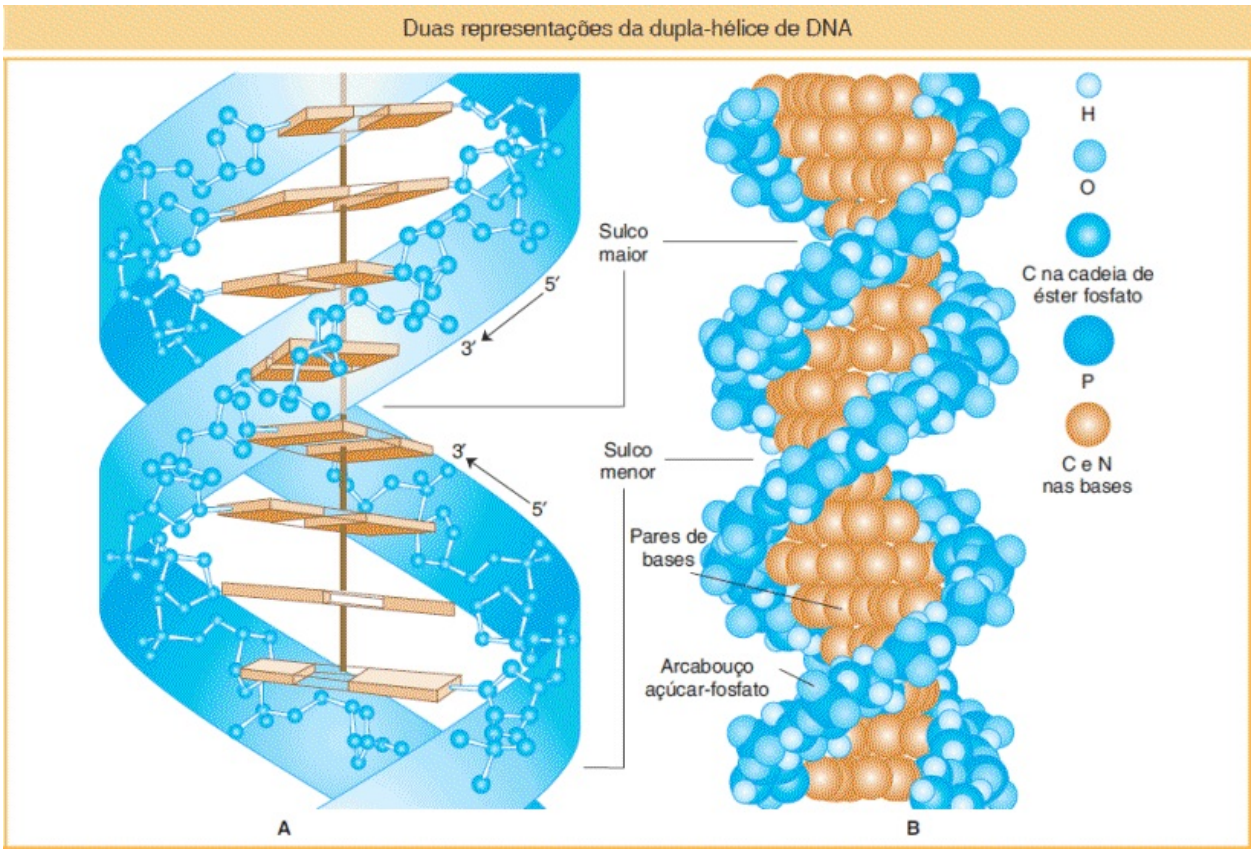


FIGURA 7.9 O diagrama do filamento (**A**) destaca o empilhamento dos pares de bases, enquanto o modelo de preenchimento de espaços (**B**) demonstra os sulcos maior e menor.

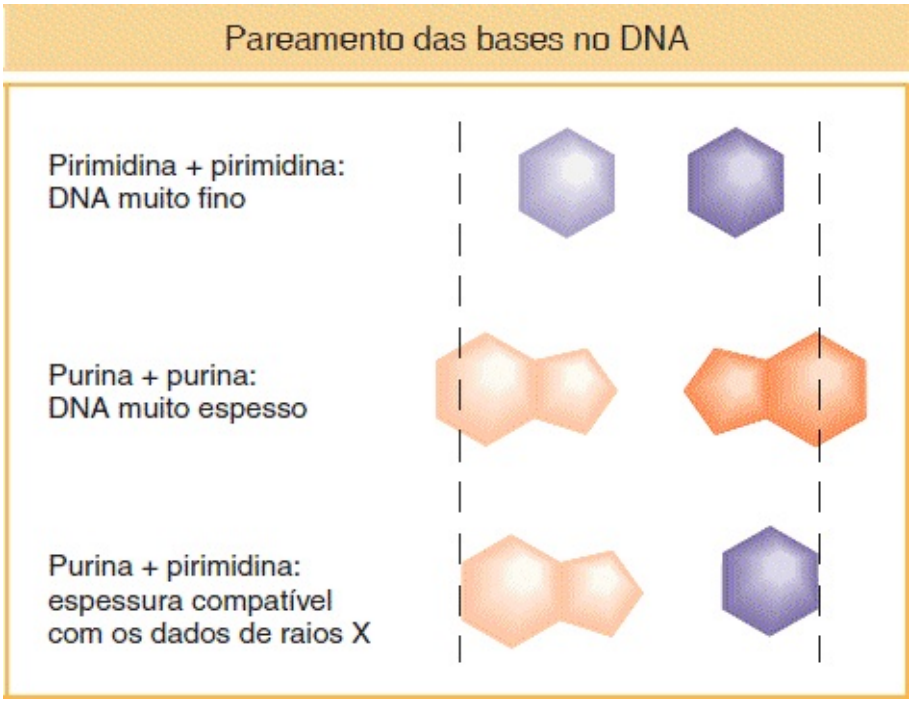


FIGURA 7.10 O pareamento das purinas com as pirimidinas explica exatamente o diâmetro da dupla-hélice de DNA determinado a partir dos dados de raios X. Aquele diâmetro é indicado pelas *linhas verticais tracejadas*.

A descoberta da estrutura do DNA por Watson e Crick é considerada por algumas pessoas a descoberta biológica mais importante do século 20 e isso fez com que eles recebessem o Prêmio Nobel com Maurice Wilkins em 1962 (Rosalind Franklin morreu de câncer em 1958 e o prêmio não é concedido postumamente). O motivo para essa descoberta ser considerada tão importante é que o modelo da dupla-hélice, além de ser consistente com os dados anteriores a respeito da estrutura do DNA, preencheu os três requisitos em relação a uma substância hereditária:

1. A estrutura da dupla-hélice sugeriu como o material genético poderia determinar a estrutura das proteínas. Talvez a *sequência* de pares de nucleotídios no DNA dite a sequência de aminoácidos na proteína especificada por aquele gene. Em outras palavras, algum tipo de **código genético** pode escrever as informações no DNA como uma sequência de nucleotídios e em seguida traduzi-las em uma linguagem diferente de sequências de aminoácidos na proteína. Como isso é realizado é o assunto do Capítulo 9.
2. Se a sequência de bases do DNA especifica a sequência de aminoácidos, então é possível a mutação por meio da substituição de um tipo de base por outro em uma ou mais posições. As mutações serão discutidas no Capítulo 16.
3. Na medida em que Watson e Crick declararam nas palavras da conclusão de seu artigo de 1953 na *Nature* que relatava a estrutura da dupla-hélice do DNA: “Não escapou da nossa atenção que o pareamento específico que postulamos sugere imediatamente um possível mecanismo de cópia em relação ao material genético.”[2] Para os geneticistas da época, o significado dessa declaração estava claro, conforme veremos na próxima seção.

7.3 Replicação semiconservativa

O mecanismo de cópia ao qual Watson e Crick fizeram referência é denominado replicação semiconservativa e está diagramado na Figura 7.11. Os arcabouços de açúcar e fosfato são representados por fitas espessas e a sequência de pares de bases é aleatória. Imaginemos que a dupla-hélice é análoga a um zíper que se abre, com início em uma extremidade. Podemos observar que, se essa analogia com o zíper é válida, a deselicoidização dos dois filamentos exporá bases únicas em cada filamento. Cada base exposta apresenta o potencial de parear com nucleotídios livres em uma solução. Tendo em vista que a estrutura do DNA impõe estritas exigências de pareamento, cada base exposta irá parear apenas com a sua **base complementar**, A com T e G com C. Portanto, cada um dos dois filamentos únicos atuará como um **modelo**, ou molde, para direcionar a montagem das bases complementares para formar novamente uma dupla-hélice idêntica à original. Presume-se que os nucleotídios recentemente adicionados advenham de um conjunto de nucleotídios livres que deve estar presente na célula.

Se esse modelo estiver correto, então cada molécula-filha deve conter uma cadeia de nucleotídios parental e uma cadeia de nucleotídios recentemente sintetizada. Entretanto, um pouco de ponderação demonstra que existem no mínimo três modos diferentes nos quais uma molécula de DNA parental pode estar relacionada com as moléculas-filhas. Esses modos hipotéticos de replicação são denominados semiconservativo (o modelo de Watson-Crick), conservativo e dispersivo (Figura 7.12). Na **replicação semiconservativa**, a dupla-hélice de cada molécula-filha de DNA contém um filamento da molécula de DNA original e um filamento recém-sintetizado. Na **replicação conservativa**, a molécula de DNA do genitor é conservada e uma única dupla hélice-filha é produzida, composta por dois filamentos recém-sintetizados. Na replicação dispersiva, *cada uma* das moléculas-filhas é composta por filamentos que contêm segmentos de *ambos* os DNA parentais e do DNA recém-sintetizado.

Experimento de Meselson-Stahl

O primeiro problema na compreensão da replicação do DNA foi descobrir se o

mecanismo da replicação era semiconservativo, conservativo ou dispersivo. Em 1958, dois jovens cientistas, Matthew Meselson e Franklin Stahl, decidiram descobrir quais dessas possibilidades descrevia corretamente a replicação do DNA. A ideia deles era possibilitar que moléculas de DNA parental contendo nucleotídios de uma densidade replicassem em um meio contendo nucleotídios de densidade diferente. Se o DNA fosse replicado de modo semiconservativo, as moléculas-filhas seriam metade antigas e metade novas e, portanto, de densidade intermediária.

Para realizar o seu experimento, Meselson e Stahl cultivaram *E. coli* em um meio contendo o isótopo pesado de nitrogênio (^{15}N), em vez do tipo leve (^{14}N) normal. Esse isótopo foi inserido nas bases nitrogenadas, que em seguida foram incorporadas em filamentos de DNA recém-sintetizados. Após muitas divisões celulares em ^{15}N, o DNA das células estava bem-marcado com o isótopo pesado. Em seguida as células foram removidas do meio com ^{15}N e colocadas em um meio com ^{14}N; após uma e duas divisões celulares, foram coletadas amostras e o DNA foi isolado de cada amostra.

Replicação semiconservativa do DNA

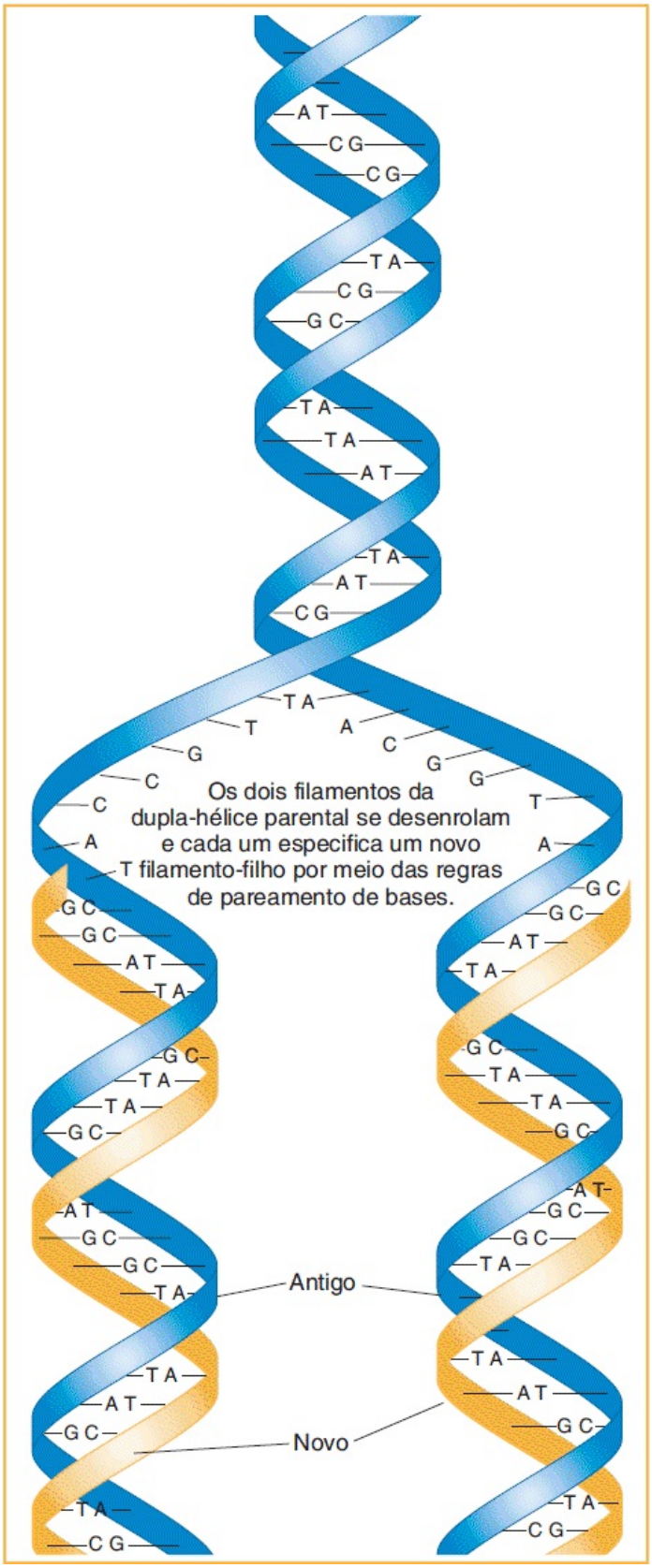


FIGURA 7.11 O modelo semiconservativo da replicação do DNA proposto por Watson e Crick tem por base a especificidade dos pares de bases ligados por pontes de hidrogênio. Os filamentos parentais, demonstrados em *azul*, atuam como moldes para a polimerização. Os filamentos recém-polimerizados, demonstrados em *amarelo*, apresentam as sequências de bases que são complementares aos seus respectivos moldes.

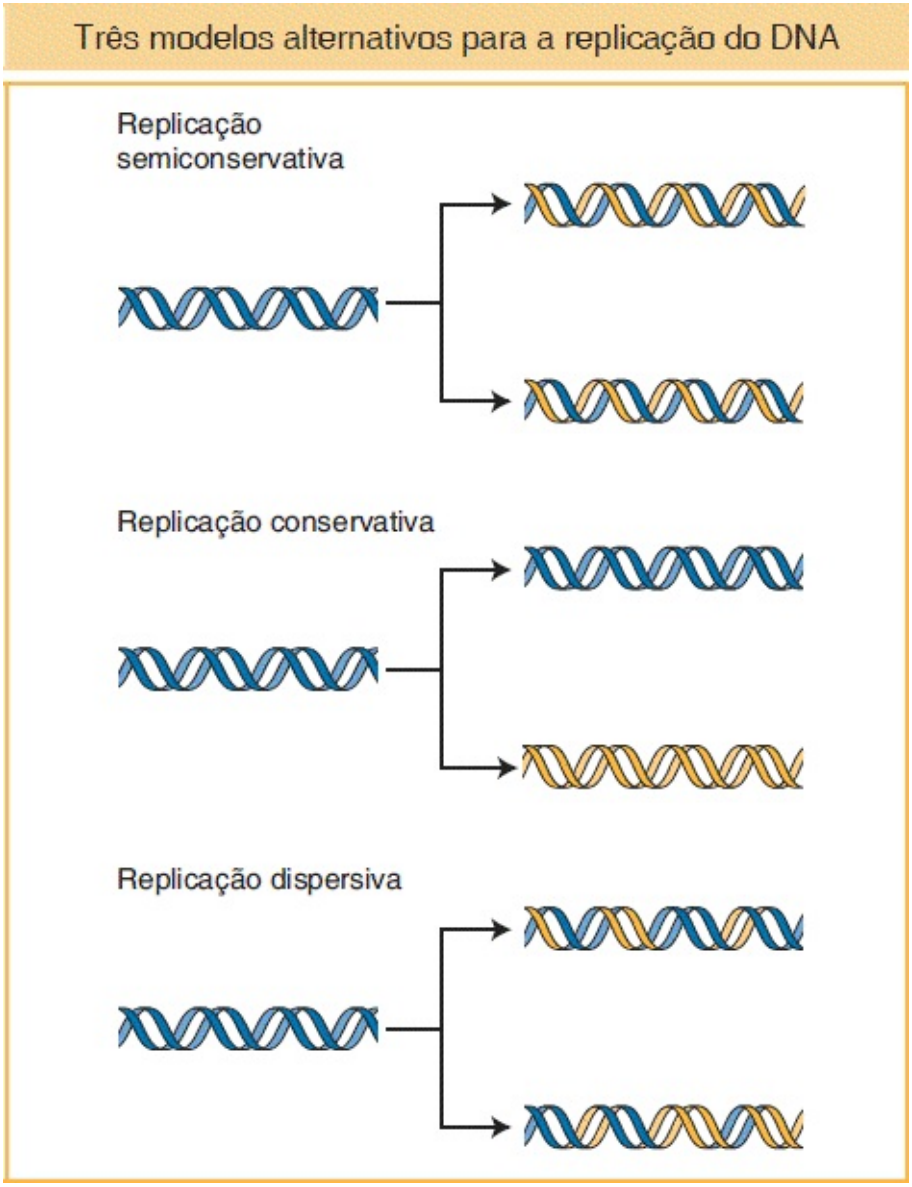


FIGURA 7.12 Dos três modelos alternativos para a replicação do DNA, o modelo da estrutura do DNA de Watson-Crick produziria o primeiro modelo (semiconservativo). As *linhas amarelas* representam os filamentos recém-sintetizados.

Meselson e Stahl foram capazes de distinguir o DNA de diferentes densidades

porque as moléculas podem ser separadas umas das outras por meio de um procedimento denominado *centrifugação em gradiente de cloreto de césio*. Se o cloreto de césio (CsCl) for centrifugado em velocidades tremendamente altas (50.000 rpm) durante muitas horas, os íons césio e cloreto tendem a ser empurrados pela força centrífuga para o fundo do tubo. Finalmente, é estabelecido um gradiente de íons no tubo, com a mais alta concentração de íons, ou densidade, no fundo. O DNA centrifugado com o cloreto de césio forma uma banda em uma posição idêntica à sua densidade no gradiente (Figura 7.13). O DNA de diferentes densidades formará bandas em locais diferentes. As células inicialmente cultivadas no isótopo pesado ^{15}N demonstraram DNA de alta densidade. Esse DNA está demonstrado em azul *à esquerda* da Figura 7.13 A. Após o cultivo dessas células no isótopo leve ^{14}N por uma geração, os pesquisadores observaram que o DNA era de densidade intermediária, demonstrando uma metade azul (^{15}N) e uma metade dourada (^{14}N) na parte intermediária da Figura 7.13 A. Observe que Meselson e Stahl continuaram o experimento durante duas gerações de *E. coli*, de modo que puderam distinguir a replicação semiconservativa da dispersiva. Após duas gerações, foram observados tanto DNA de densidade intermediária quanto de densidade baixa (*à direita* da Figura 7.13 A), precisamente como previsto pelo modelo de replicação semiconservativa de Watson e Crick.

> CONCEITO-CHAVE O DNA é replicado por meio da deselicoidização dos dois filamentos da dupla-hélice e da construção de um novo filamento complementar em cada um dos filamentos separados da dupla-hélice original.

Forquilha de replicação

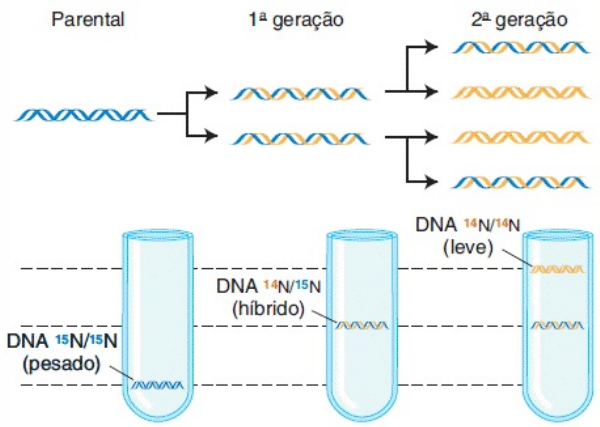
Outra previsão do modelo de replicação do DNA de Watson-Crick é que um zíper de replicação, ou *forquilha*, será observado na molécula de DNA durante a replicação. Essa forquilha é o local no qual a dupla-hélice é desenrolada para produzir os dois filamentos únicos que atuam como moldes para a cópia. Em

1963, John Cairns testou essa previsão ao possibilitar a replicação do DNA em células bacterianas para incorporar a timidina tritiada ([^{3}H]timidina) — o nucleotídio da timidina marcado com um isótopo de hidrogênio radioativo denominado trítio. Teoricamente, cada molécula-filha recentemente sintetizada deve conter um filamento radioativo (“quente”) (com ^{3}H) e outro filamento não radioativo (“frio”). Após intervalos variados e quantidades variadas de ciclos de replicação em um meio “quente”, Cairns lisou cuidadosamente as bactérias e possibilitou que os conteúdos celulares assentassem em lâminas de microscopia eletrônica. Finalmente, Cairns cobriu as lâminas com emulsão fotográfica e a expôs ao escuro durante 2 meses. Esse procedimento, denominado autorradiografia, possibilitou que Cairns desenvolvesse uma imagem da localização de ^{3}H no material celular. Na medida em que o ^{3}H decai, ele emite uma partícula beta (um elétron energético). A emulsão fotográfica detecta uma reação química que ocorre sempre que uma partícula beta atinge a emulsão. Em seguida, a emulsão pode ser revelada como uma impressão fotográfica, de modo que o rastro de emissão de partículas beta aparece como pontos ou grãos pretos.

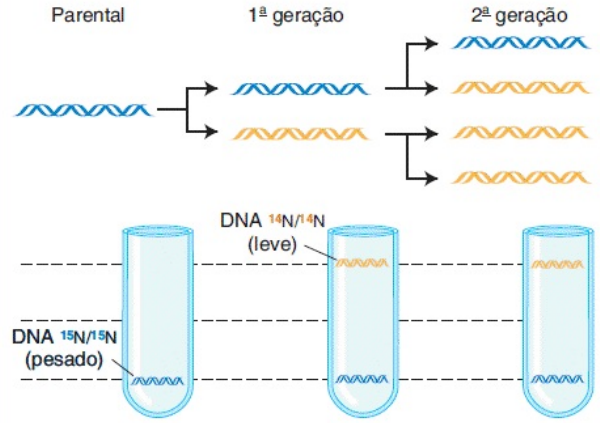
Após um ciclo de replicação em [^{3}H]timidina, apareceu um anel de pontos na autorradiografia. Cairns interpretou esse anel como um filamento radioativo recentemente formado em uma molécula-filha de DNA circular, conforme demonstrado na Figura 7.14 A. Portanto, estava aparente que o cromossomo bacteriano é circular — um fato que também surgiu a partir da análise genética descrita anteriormente (Capítulo 5). No segundo ciclo de replicação, as forquilhas previstas pelo modelo de fato foram observadas. Além disso, a densidade dos grãos nos três segmentos foi tal que a interpretação demonstrada na Figura 7.14 B pode ser realizada: a curva espessa de pontos que corta o interior do círculo de DNA seria o filamento-filho recém-sintetizado, dessa vez composto por *dois* filamentos radioativos. Cairns observou todos os tamanhos desses padrões autorradiográficos em formato de lua, correspondendo à movimentação progressiva das forquilhas de replicação, ao redor do anel.

O DNA é copiado por meio
da replicação semiconservativa

A. Previsões do modelo semiconservativo



B. Previsões do modelo conservativo



C. Previsões do modelo dispersivo

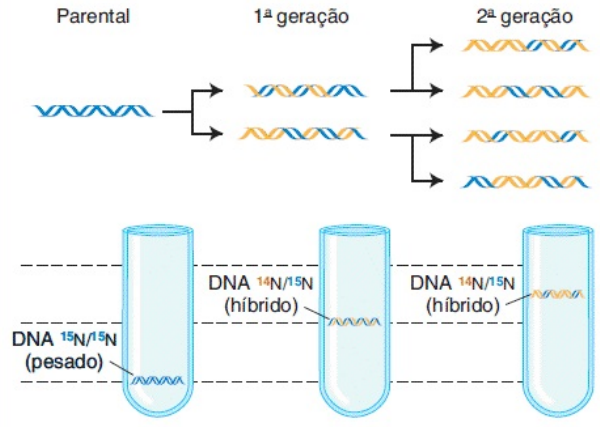


FIGURA 7.13 O experimento de Meselson-Stahl demonstra que o DNA é copiado por meio da replicação semiconservativa. O DNA centrifugado em um gradiente de cloreto de césio (CsCl) formará bandas de acordo com a sua densidade. **A.** Quando as células cultivadas em ^{15}N são transferidas para um meio com ^{14}N, a primeira geração produz uma única banda de DNA intermediária e a segunda geração produz duas bandas: uma intermediária e uma leve. Este resultado corresponde às previsões do modelo semiconservativo de replicação do DNA. **B** e **C.** Os resultados previstos em relação à replicação conservativa e replicação dispersiva, demonstrados aqui, *não* foram observados.

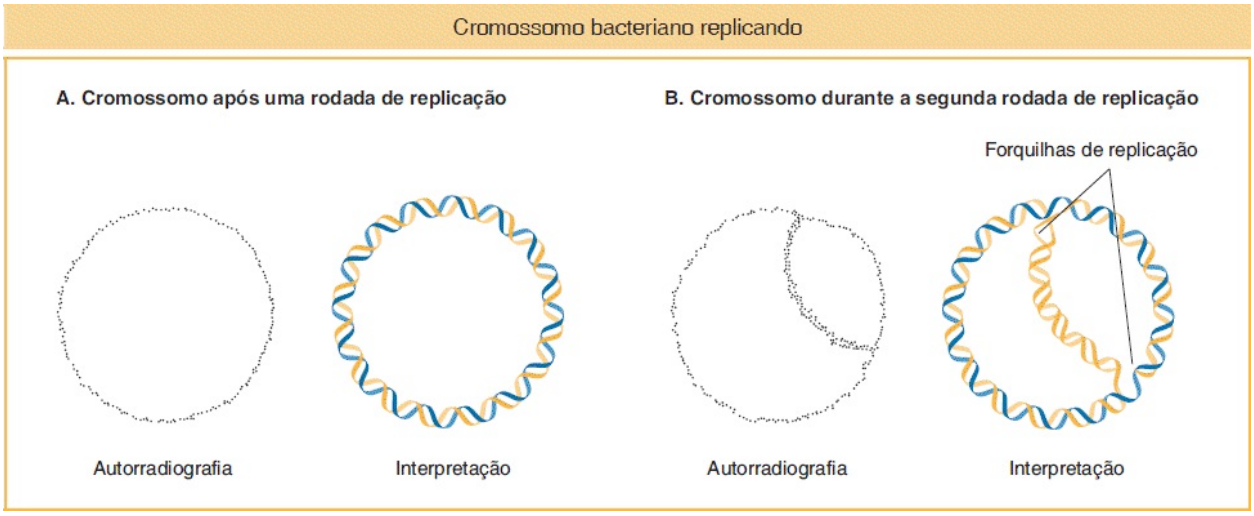


FIGURA 7.14 Um cromossomo bacteriano replicando apresenta duas forquilhas de replicação. **A.** *Esquerda*: autorradiografia de um cromossomo bacteriano após uma replicação em timidina tritiada. De acordo com o modelo de replicação semiconservativa, um dos dois filamentos deve ser radioativo. *Direita*: interpretação da autorradiografia. A hélice *amarela* representa o filamento tritiado. **B.** *Esquerda*: autorradiografia de um cromossomo bacteriano na segunda rodada da replicação em timidina tritiada (^{3}H). Nesta molécula, a dupla-hélice recentemente replicada que cruza o círculo pode ser composta por dois filamentos radioativos (se o filamento parental for o filamento radioativo). *Direita*: a espessura dupla do traçado radioativo no autorradiograma aparenta confirmar a interpretação demonstrada aqui.

DNA polimerases

Um problema confrontado pelos cientistas era compreender justamente como as bases são trazidas até o molde da dupla-hélice. Embora os cientistas suspeitassem que enzimas desempenhassem um papel, essa possibilidade não foi comprovada até 1959, quando Arthur Kornberg isolou a DNA polimerase de *E. coli* e demonstrou sua atividade enzimática *in vitro*. Essa enzima adiciona desoxirribonucleotídios na extremidade 3′ de uma cadeia de nucleotídios em

crescimento, utilizando como molde um filamento único de DNA que foi exposto pela deselicoidização localizada da dupla-hélice (Figura 7.15). Os substratos para a DNA polimerase são as formas trifosfato dos desoxirribonucleotídios, dATP, dGTP, dCTP e dTTP. A adição de cada base ao polímero em crescimento é acompanhada pela remoção de dois dos três fosfatos na forma de pirofosfato (PP_i). A energia produzida pela clivagem dessa ligação de alta energia e a subsequente hidrólise do pirofosfato em duas moléculas de fosfato inorgânico auxiliam no direcionamento do processo endergônico de construção de um polímero de DNA.

Atualmente sabe-se que existem cinco DNA polimerases na *E. coli*. A primeira enzima que Kornberg purificou atualmente é denominada DNA polimerase I, ou pol I. Essa enzima apresenta três atividades, que aparentam estar localizadas em diferentes partes da molécula:

1. Uma atividade de polimerase, que catalisa o crescimento da cadeia no sentido 5′ para 3′.
2. Uma atividade de exonuclease 3′ para 5′, que remove bases incorretamente pareadas.
3. Uma atividade de exonuclease 5′ para 3′, que degrada filamentos únicos de DNA ou RNA.

Retornaremos à significância das duas atividades de exonuclease posteriormente neste capítulo.

Embora a pol I atue na replicação do DNA (ver próxima seção), alguns cientistas suspeitavam de que ela não fosse responsável pela maior parte da síntese do DNA, tendo em vista que ela era muito lenta (aproximadamente 20 nucleotídios/s) e muito abundante (aproximadamente 400 moléculas/célula), além de se dissociar do DNA após a incorporação de apenas 20 a 50 nucleotídios. Em 1969, John Cairns e Paula DeLucia resolveram essa questão quando demonstraram que uma linhagem de *E. coli* com uma mutação no gene que codifica a DNA pol I ainda era capaz de crescer normalmente e replicar o seu DNA. Eles concluíram que outra DNA polimerase, atualmente denominada pol III,

catalisa a síntese de DNA na forquilha de replicação.

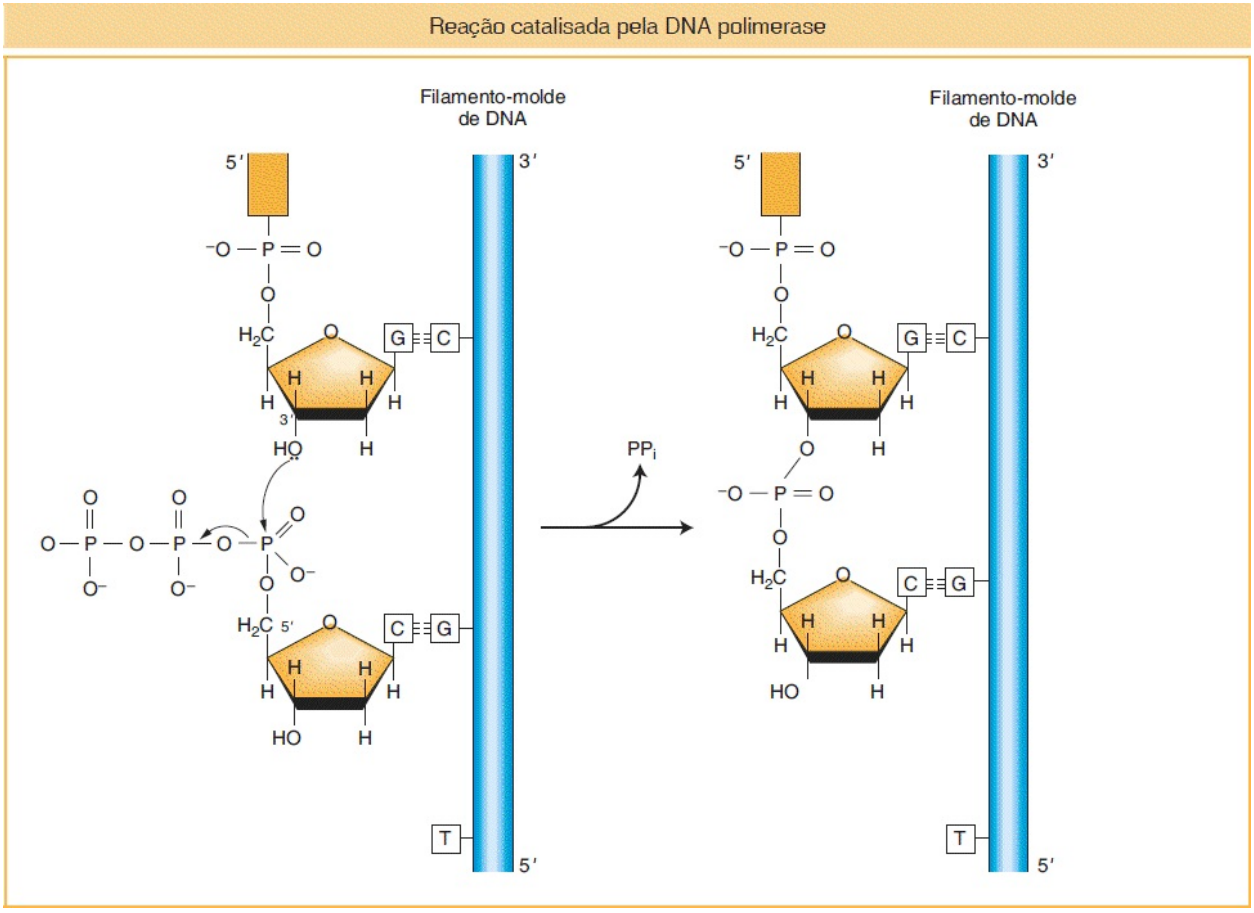


FIGURA 7.15 A DNA polimerase catalisa a reação de alongamento da cadeia. A energia para a reação advém da quebra da ligação fosfato de alta energia do substrato trifosfato.

7.4 Visão geral da replicação do DNA

Na medida em que a DNA pol III avança, a dupla-hélice é continuamente desenrolada à frente da enzima para expor comprimentos adicionais de filamentos de DNA simples que atuarão como moldes (Figura 7.16). A DNA pol III atua como a **forquilha de replicação**, a zona na qual a dupla-hélice está desenrolando. Entretanto, tendo em vista que a DNA polimerase sempre adiciona nucleotídios na *extremidade* 3′ *em crescimento*, apenas um dos dois filamentos antiparalelos pode atuar como molde para a replicação na direção da forquilha de replicação. Em relação a esse filamento, a síntese pode ocorrer de modo contínuo e suave na

direção da forquilha; o novo filamento no sentido líder, é denominado **filamento contínuo** (no sentido líder, ***leading***).

A síntese do outro filamento também ocorre na extremidade em crescimento 3′, mas essa síntese ocorre no sentido "errado", tendo em vista que, em relação a esse filamento, o sentido 5′ para 3′ da síntese está longe da forquilha de replicação (ver Figura 7.16). Conforme veremos, a natureza do maquinário de replicação requer que a síntese de ambos os filamentos ocorra na região da forquilha de replicação. Portanto, a síntese que se afasta da forquilha de crescimento não pode prosseguir por muito tempo. Ela tem de ocorrer em segmentos curtos: a polimerase sintetiza um segmento, em seguida se movimenta de volta para a extremidade 5′ do segmento, onde a forquilha em crescimento expôs o novo molde e inicia o processo novamente. Esses trechos curtos (1.000 a 2.000 nucleotídios) de DNA recém-sintetizado são denominados **fragmentos de Okazaki**.

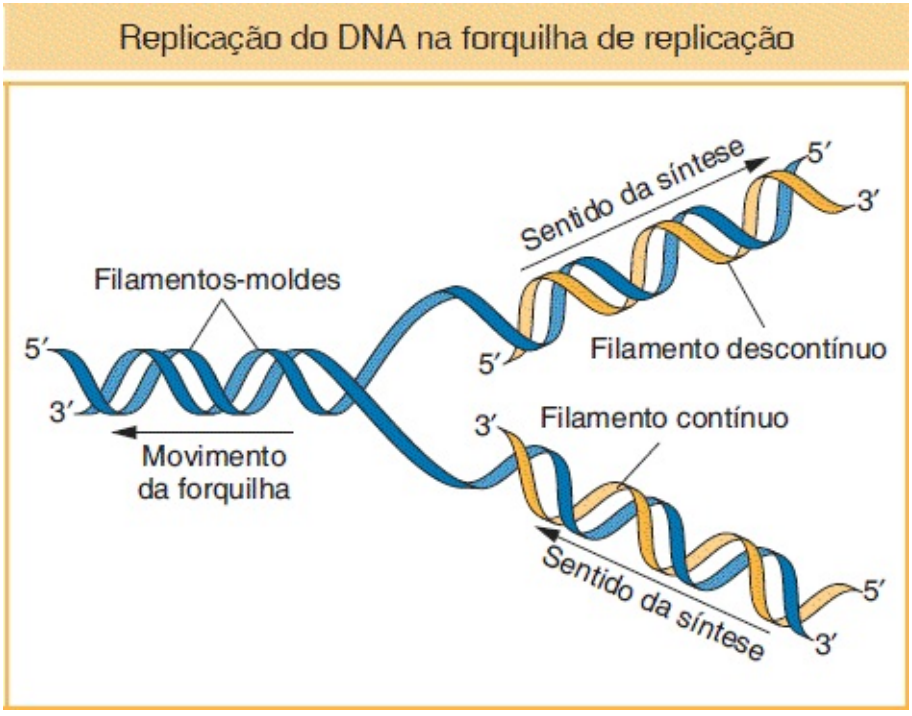


FIGURA 7.16 A forquilha de replicação se movimenta na síntese de DNA na medida em que a dupla-hélice é continuamente desenrolada. A síntese do filamento contínuo (*leading*) pode prosseguir suavemente sem interrupções na direção do movimento da forquilha de replicação, mas a síntese do filamento descontínuo (*lagging*) deve prosseguir na direção oposta, para longe da forquilha de replicação.

Outro problema na replicação do DNA surge em virtude de a DNA polimerase conseguir estender uma cadeia, mas não iniciar uma cadeia. Portanto, a síntese do filamento contínuo e de cada fragmento de Okazaki tem de ser iniciada por um ***primer***, ou uma cadeia curta de nucleotídios, que se liga ao filamento-molde para formar um segmento de ácido nucleico dúplex. O *primer* na replicação do DNA pode ser observado na Figura 7.17. Os *primers* são sintetizados por um conjunto de proteínas denominado **primossomo**, do qual um componente central é uma enzima denominada **primase**, um tipo de RNA polimerase. A primase sintetiza um trecho curto (aproximadamente 8 a 12 nucleotídios) de RNA complementar a uma região específica do cromossomo. No filamento contínuo, é necessário apenas um *primer* inicial, tendo em vista que, após a ação do *primer* inicial, o filamento de DNA em crescimento atua como *primer* para a adição contínua. Entretanto, no filamento descontínuo, cada fragmento de Okazaki necessita de seu próprio *primer*. A cadeia de RNA que compõe o *primer* em seguida é estendida como uma cadeia de DNA pela DNA pol III.

Uma DNA polimerase diferente, pol I, remove os *primers* de RNA com sua atividade de exonuclease 5′ para 3′ e preenche as lacunas com sua atividade de polimerase 3′ para 5′. Conforme mencionado anteriormente, a pol I é a enzima originalmente purificada por Kornberg. Outra enzima, a **DNA ligase**, une a extremidade 3′ do DNA que preencheu a lacuna à extremidade 5′ do fragmento de Okazaki *dowstream*. O novo filamento assim formado é denominado **filamento descontínuo (*lagging*)**. A DNA ligase une os fragmentos do DNA ao catalisar a formação de uma ligação fosfodiéster entre a extremidade 5′-fosfato de um fragmento e o grupo 3′-OH adjacente de outro fragmento.

Um marco da replicação do DNA é a sua precisão, também denominada fidelidade: em geral, é inserido menos de um erro por 10^{10} nucleotídios. Parte do motivo da precisão da replicação do DNA é que ambas a DNA pol I e a DNA pol III possuem atividade de exonuclease 3′ para 5′, que atua como uma função de “revisão” por meio da excisão de bases incorretamente inseridas.

Em virtude da importância da revisão, veremos mais de perto como ela atua. Um par de bases incorretamente pareado ocorre quando a atividade da polimerase 5′ para 3′ insere, por exemplo, uma A em vez de uma G para parear

com uma C. A adição de uma base incorreta com frequência ocorre em virtude de um processo denominado **tautomerização**. Cada uma das bases no DNA pode se apresentar em uma de diversas formas, denominadas tautômeros, que são isômeros que diferem nas posições de seus átomos e nas ligações entre os átomos. As formas estão em equilíbrio. A forma **ceto** de cada base normalmente está presente no DNA, mas em casos raros, uma base pode ser alterada para a forma **imino** ou **enol**. As formas imino e enol podem parear com a base errada, formando um par errôneo (Figura 7.18). Quando uma C é alterada para a sua forma rara imino, a polimerase adiciona uma A em vez de uma G (Figura 7.19). Felizmente, tal erro normalmente é detectado e removido pela atividade de exonuclease 3′ para 5′. Após a remoção da base incorreta, a polimerase tem outra chance de adicionar a base G complementar correta.

Conforme seria esperado, linhagens mutantes sem uma exonuclease 3′ para 5′ funcional apresentam uma taxa mais alta de mutação. Além disso, tendo em vista que a primase não apresenta uma função de revisão, o *primer* de RNA apresenta maior probabilidade de conter erros do que o DNA. A necessidade de manter a alta fidelidade da replicação é um motivo pelo qual os *primers* de RNA nas extremidades dos fragmentos de Okazaki precisam ser removidos e substituídos por DNA. Apenas após a remoção do *primer* de RNA a DNA pol I catalisa a síntese do DNA para substituir o *primer*. O assunto do reparo do DNA será comentado em detalhes no Capítulo 16.

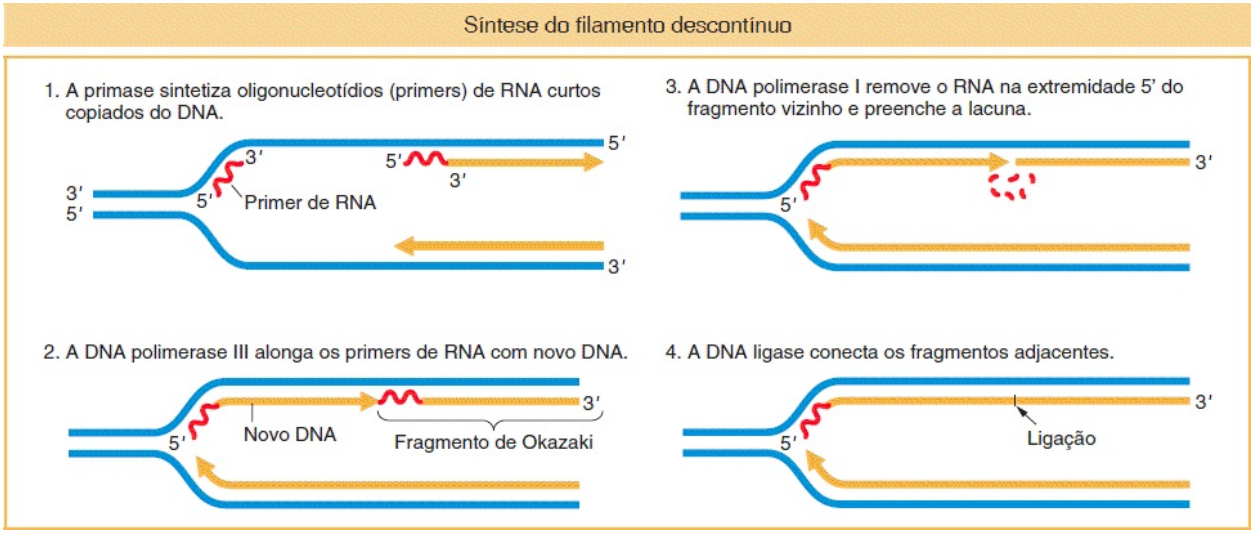


FIGURA 7.17 Etapas na síntese do filamento descontínuo (*lagging*). A síntese de DNA prossegue por meio da síntese contínua no filamento *leading* e da síntese descontínua no filamento *lagging*.

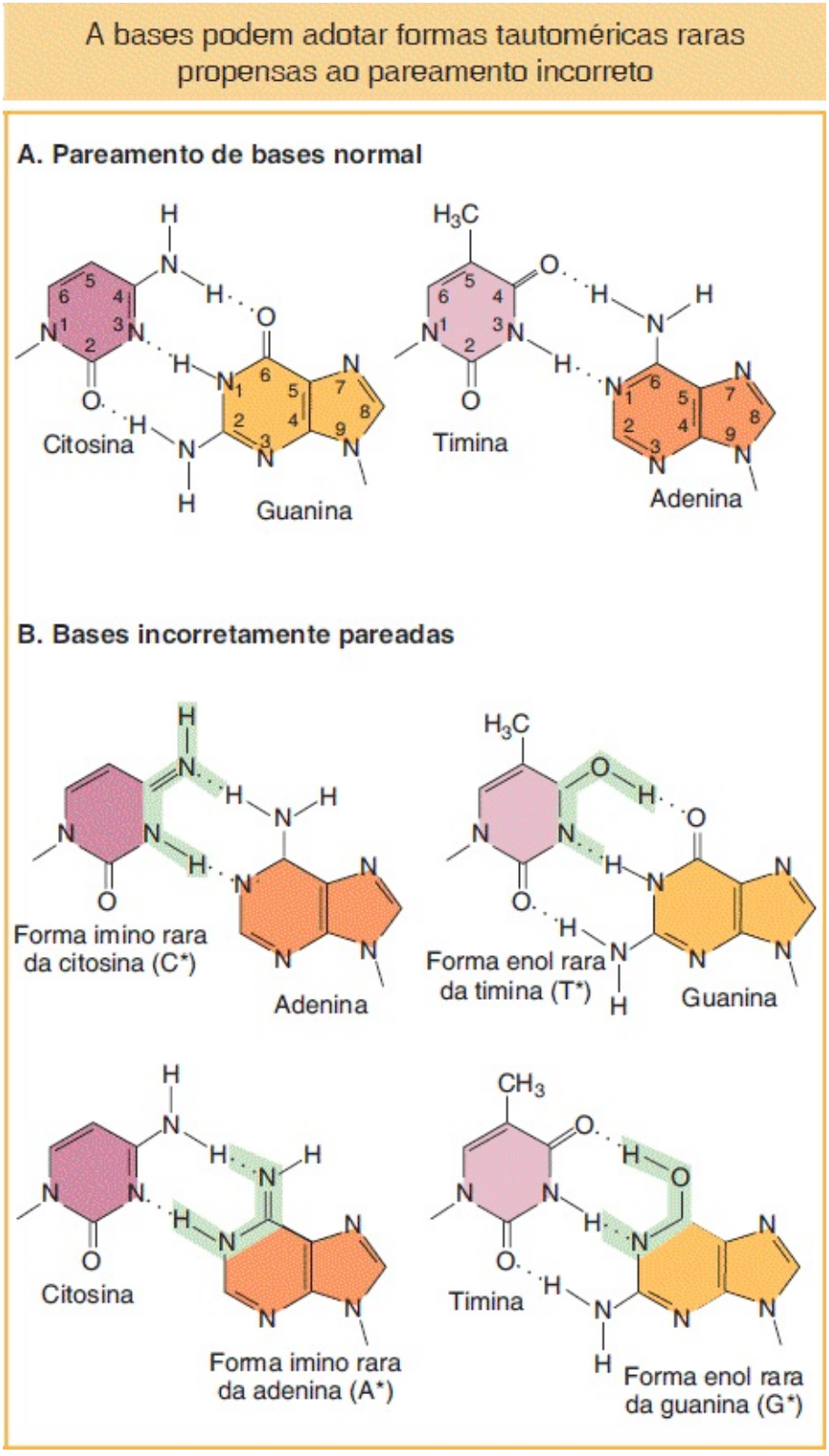


FIGURA 7.18 Pareamento de bases normal, em comparação com o de bases incorretamente pareadas. **A.** Pareamento entre as formas normais (ceto) das bases. **B.** Formas tautoméricas raras de bases resultam em pareamentos incorretos.

CONCEITO-CHAVE A replicação do DNA ocorre na forquilha de replicação, onde a dupla-hélice está desenrolando e os dois filamentos estão se separando. A replicação do DNA prossegue continuamente na direção da forquilha de replicação em deselicoidização no filamento contínuo. O DNA é sintetizado em segmentos curtos, na direção para longe da forquilha de replicação, no filamento descontínuo. A DNA polimerase necessita que um *primer*, ou uma cadeia curta de nucleotídios, já esteja posicionado para iniciar a síntese.

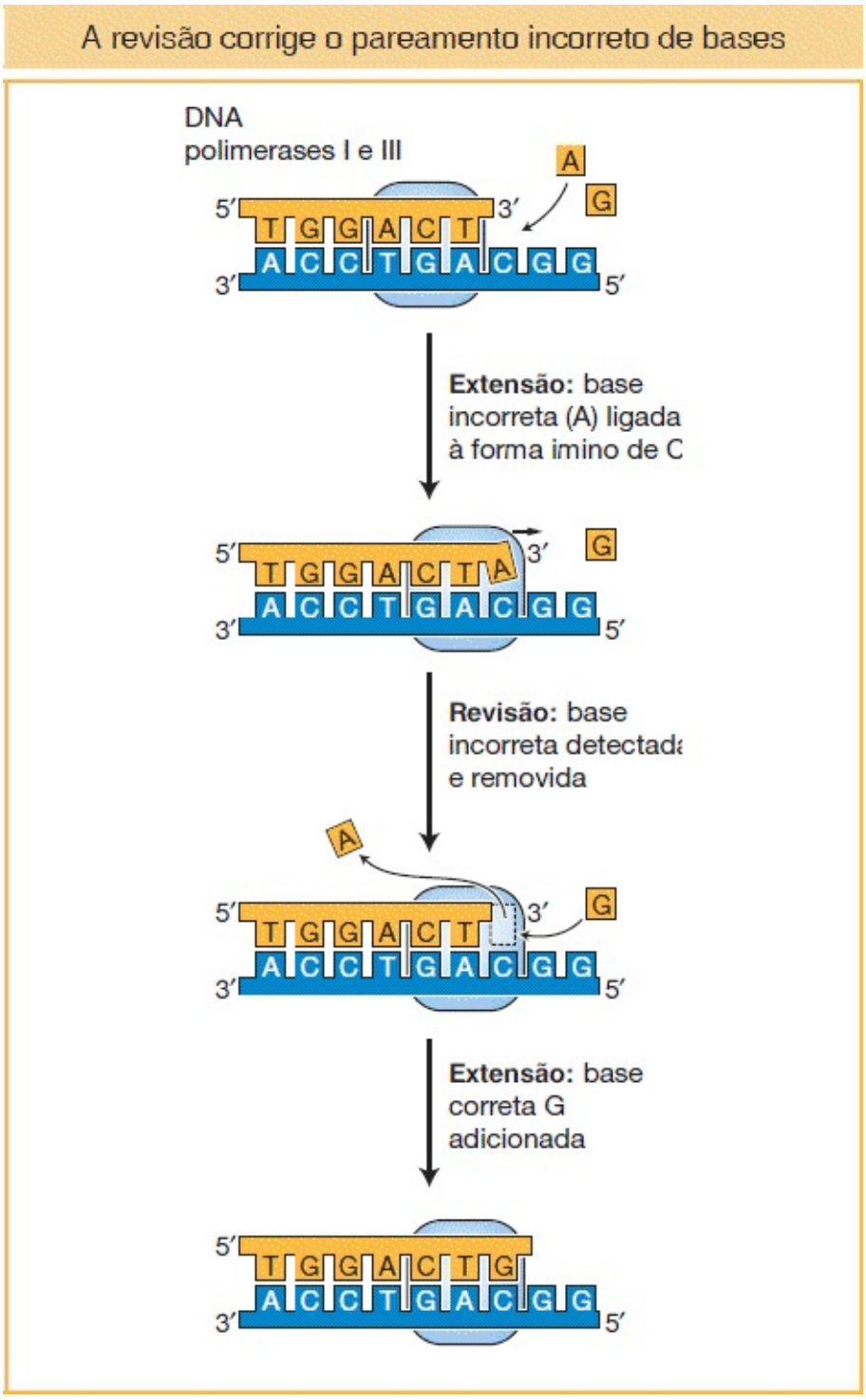


FIGURA 7.19 A DNA polimerase remove o pareamento incorreto A-C com a utilização da sua atividade de exonuclease 3′ para 5′.

7.5 Replissomo | Uma máquina de replicação notável

Outro marco da replicação do DNA é a velocidade. O tempo necessário para que a *E. coli* replique o seu cromossomo é tão breve quanto 40 min. Portanto, seu genoma de aproximadamente 5 milhões de pares de bases deve ser copiado a uma velocidade de aproximadamente *2.000 nucleotídios por segundo*. A partir do experimento de Cairns, sabemos que a *E. coli* utiliza apenas duas forquilhas de replicação para copiar o seu genoma inteiro. Portanto, cada forquilha tem de ser capaz de se movimentar a uma velocidade de até *1.000 nucleotídios por segundo*. O que é notável a respeito de todo o processo de replicação do DNA é que ele não sacrifica a velocidade pela precisão. Como ele consegue manter ambas, a velocidade e a precisão, tendo em vista a complexidade das reações na forquilha de replicação? A resposta é que a DNA polimerase é parte de um grande complexo "nucleoproteico" que coordena as atividades na forquilha de replicação. Esse complexo, denominado **replissomo**, é um exemplo de uma "máquina molecular". Você encontrará outros exemplos nos capítulos posteriores. A descoberta de que as principais funções da célula — replicação, transcrição e tradução, por exemplo — são realizadas por grandes complexos de multissubunidades alterou o modo como pensamos a respeito das células. Para começar a compreender o motivo, vejamos o replissomo mais de perto.

Alguns dos componentes do replissomo na *E. coli* em interação estão demonstrados na Figura 7.20. Na forquilha de replicação, o centro catalítico da DNA pol III for parte de um complexo muito maior, denominado **holoenzima pol III**, que é composto por dois centros catalíticos e muitas proteínas acessórias. Um dos centros catalíticos lida com a síntese do filamento contínuo, enquanto o outro lida com a síntese do filamento descontínuo. Algumas das proteínas acessórias (não visíveis na Figura 7.20) formam uma conexão entre os dois centros catalíticos, coordenando, assim, a síntese dos filamentos contínuo e descontínuo. O filamento descontínuo está demonstrado fazendo uma alça, de modo que o replissomo consegue coordenar a síntese de ambos os filamentos e se movimentar para a forquilha de replicação. Uma proteína acessória importante, denominada grampo β, circunda o DNA como uma rosquinha e mantém a pol III ligada à molécula de DNA. Portanto, a pol III é transformada de uma enzima que pode

adicionar apenas 10 nucleotídios antes de sair do molde (denominada **enzima distributiva**) em uma enzima que permanece na forquilha em movimento e que adiciona dezenas de milhares de nucleotídios (uma **enzima processiva**). Resumidamente, por meio da ação de proteínas acessórias, a síntese de ambos os filamentos contínuo e descontínuo é rápida e altamente coordenada.

Proteínas em atuação na forquilha de replicação

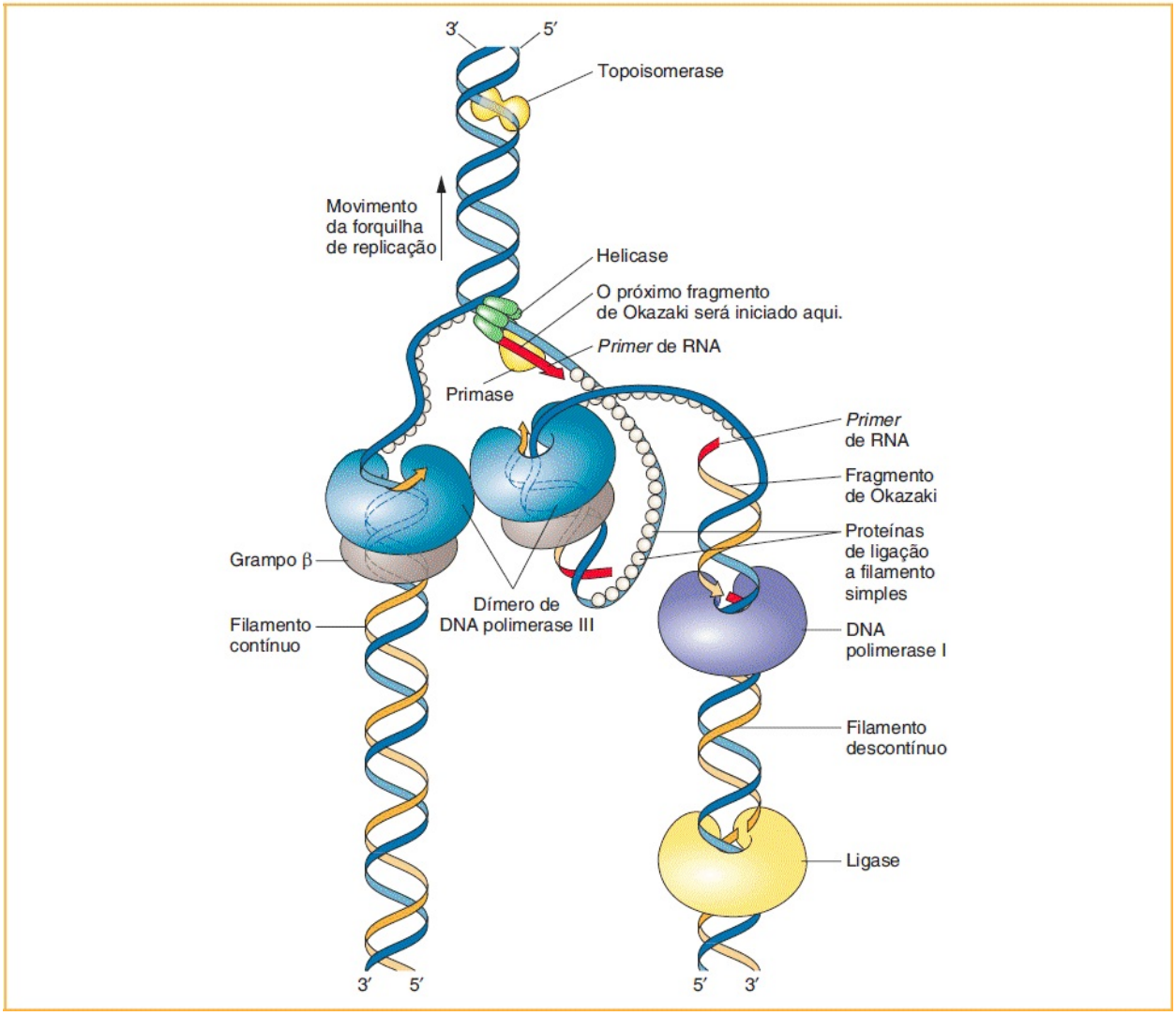


FIGURA 7.20 O replissomo e as proteínas acessórias realizam uma diversidade de etapas na forquilha de replicação. A topoisomerase e a helicase desenrolam e abrem a dupla-hélice na preparação para a replicação do DNA. Quando a dupla-hélice foi desenrolada, proteínas de ligação a filamento simples evitam que a dupla-hélice se forme novamente. A ilustração é uma representação do assim denominado modelo em trombone (denominado em virtude da sua semelhança com um trombone em virtude da alça do filamento descontínuo), demonstrando como se imagina que os dois centros catalíticos do replissomo interajam para coordenar os

diversos eventos de replicação dos filamentos contínuo e descontínuo.

Observe que a primase, a enzima que sintetiza o *primer* de RNA, não está tocando na proteína grampo. Portanto, a primase atua como uma enzima distributiva — ela adiciona apenas alguns ribonucleotídios antes de se dissociar do molde. Esse modo de ação faz sentido, tendo em vista que o *primer* precisa ser apenas longo o suficiente para formar um ponto inicial dúplex adequado para a DNA pol III.

Deselicoidização da dupla-hélice

Quando a dupla-hélice foi proposta em 1953, uma objeção importante era que a replicação de uma referida estrutura exigiria a deselicoidização da dupla-hélice na forquilha de replicação e a quebra das ligações de hidrogênio que mantêm os filamentos unidos. Como o DNA poderia ser desenrolado tão rapidamente e, ainda que pudesse, como não helicoidizar excessivamente o DNA atrás da forquilha e o tornar extremamente emaranhado? Atualmente sabemos que o replissomo contém duas classes de proteínas que abrem a hélice e evitam a helicoidização excessiva: elas são as **helicases** e as **topoisomerases**, respectivamente.

As helicases são enzimas que rompem as ligações de hidrogênio que mantêm os dois filamentos da dupla-hélice juntos. Assim como a proteína grampo, a helicase se adéqua como uma rosquinha ao redor do DNA; a partir dessa posição, ela rapidamente desenrola a dupla-hélice à frente da síntese do DNA. O DNA desenrolado é estabilizado por meio de **proteínas de ligação a filamento simples (SSB**, do inglês, *single-strand-binding***)**, que se ligam ao DNA unifilamentar e evitam a reestruturação do dúplex.

O DNA circular pode ser torcido e espiralado, de modo muito semelhante às espirais extras que podem ser introduzidas em um elástico torcido. O desenrolar da forquilha de replicação pelas helicases causa torção extra em outras regiões e são formadas superespirais para liberar a tensão da torção extra. Tanto as torções quanto as superespirais têm de ser removidas para possibilitar que a replicação continue. Essa super-helicoidização pode ser criada ou relaxada por enzimas

denominadas topoisomerases, das quais um exemplo é a DNA girase (Figura 7.21). As topoisomerases relaxam o DNA super-helicoidizado por meio da quebra de um único filamento de DNA ou de ambos os filamentos, o que possibilita que o DNA gire tornando-se uma molécula relaxada. As topoisomerases encerram reunindo os filamentos da molécula de DNA agora relaxada.

> CONCEITO-CHAVE Uma máquina molecular denominada replissomo realiza a síntese do DNA. Ela inclui duas unidades de DNA polimerase para lidar com a síntese em cada filamento e coordena a atividade das proteínas acessórias necessárias para o *priming*, a deselicoidização da dupla-hélice e a estabilização dos filamentos separados.

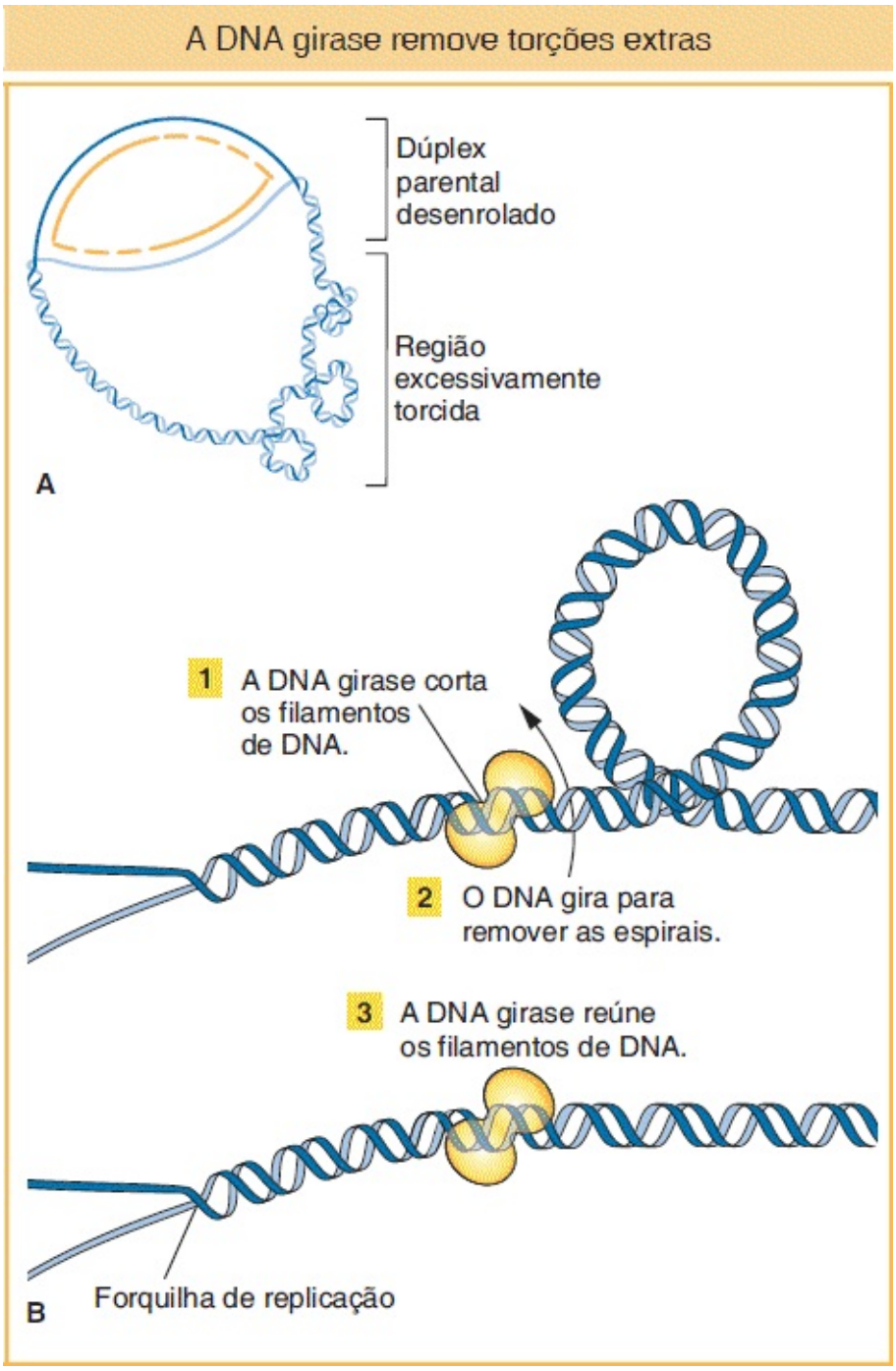


FIGURA 7.21 A DNA girase, uma topoisomerase, remove torções extras durante a replicação. **A.** Regiões extratorcidas (positivamente super-helicoidizadas) se acumulam à frente da forquilha na medida em que os filamentos parentais se separam para a replicação. **B.** Uma topoisomerase, tal como a DNA girase, remove estas regiões por meio de corte nos filamentos de DNA, possibilitando que eles girem, e, em seguida, reúne os filamentos.

Montagem do replissomo | Início da replicação

A montagem do replissomo é um processo ordenado que tem início em locais precisos no cromossomo (denominados **origens**) e que ocorre apenas em determinadas ocasiões na vida da célula. A replicação da *E. coli* tem início a partir de uma origem fixa (um *locus* denominado *oriC*) e em seguida prossegue em ambas as direções (com forquilhas movendo-se em ambas as extremidades, conforme demonstrado na Figura 7.14), até que as forquilhas se fundam. A Figura 7.22 demonstra o processo de montagem do replissomo. A primeira etapa é a ligação de uma proteína denominada DnaA a uma sequência de 13 pares de bases (pb) específica (denominada "DnaA box"), que é repetida cinco vezes em *oriC*. Em resposta à ligação da DnaA, a origem é deselicoidizada em um grupo de nucleotídios A e T. Relembre que os pares de bases AT são mantidos juntos por apenas duas ligações de hidrogênio, enquanto os pares de bases GC são mantidos juntos por três. Portanto, é mais fácil separar (dissociar) a dupla-hélice em trechos de DNA que são enriquecidos com bases A e T.

Iniciação da replicação procariótica

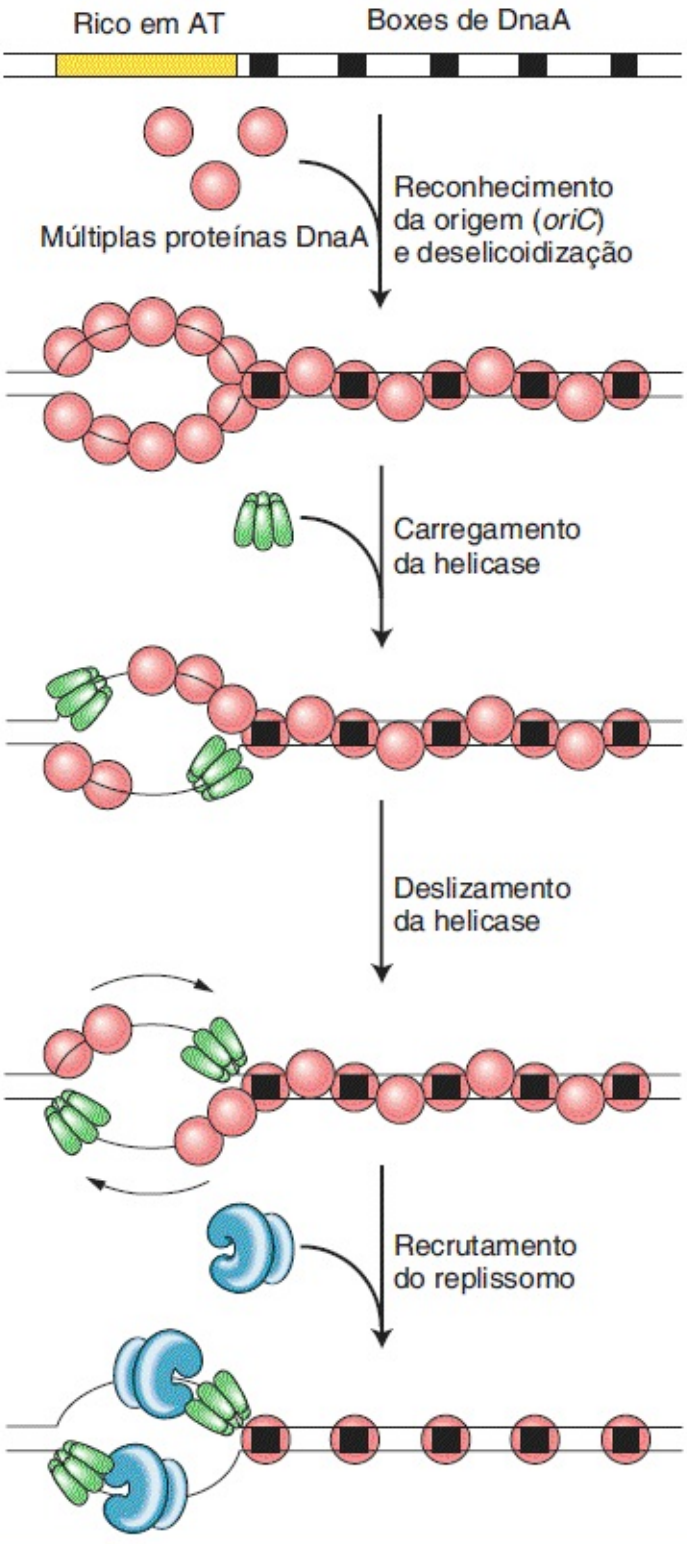


FIGURA 7.22 A síntese de DNA é iniciada nas origens de replicação em procariotos. As proteínas se ligam à origem (*oriC*), onde separam os dois filamentos da dupla-hélice e recrutam componentes do replissomo para as duas forquilhas de replicação.

Após o início da deselicoidização, proteínas DnaA adicionais se ligam às regiões de filamentos simples recém-deselicoidizadas. Com a DnaA revestindo a origem, duas helicases (proteína DnaB) agora se ligam e deslizam de 5′ para 3′ a fim de iniciar a abertura da hélice na forquilha de replicação. A primase e a holoenzima DNA pol III são recrutadas para a forquilha de replicação por meio de interações proteína-proteína e a síntese do DNA tem início. Você pode estar pensando o motivo de a DnaA não estar presente na Figura 7.20, que demonstra o replissomo. A resposta é que, embora ela seja necessária para a montagem do replissomo, ela não é parte da máquina de replicação. Em vez disso, sua função é posicionar o replissomo no local correto no cromossomo circular para a iniciação da replicação.

7.6 Replicação em organismos eucariotos

A replicação do DNA em procariotos e eucariotos utiliza um mecanismo semiconservativo e emprega a síntese de filamentos contínuo e descontínuo. Por esse motivo, não deve ser uma surpresa que os componentes do replissomo procariótico e aqueles do replissomo eucariótico sejam bastante semelhantes. Entretanto, na medida em que os organismos aumentam em complexidade, o número de componentes do replissomo também aumenta.

Origens da replicação eucariótica

Bactérias tais como *E. coli* normalmente completam um ciclo de replicação e divisão em 20 a 40 minutos, mas, em eucariotos, o ciclo pode variar de 1,4 hora em leveduras a 24 horas em células animais cultivadas e pode durar de 100 a 200 horas em algumas células. Os eucariotos têm de resolver o problema da coordenação da replicação de mais de um cromossomo.

Para compreender as origens da replicação eucariótica, primeiramente

voltaremos a nossa atenção para a levedura, um eucarioto simples. Muitas proteínas eucarióticas que atuam nas origens de replicação foram identificadas pela primeira vez em leveduras, em virtude da facilidade da análise genética em pesquisas com leveduras (ver Organismo-modelo, Levedura, no Capítulo 12). As origens de replicação em leveduras são muito semelhantes à *oriC* em *E. coli*. As origens de 100 a 200 pb apresentam uma sequência de DNA conservada que inclui uma região rica em AT que se dissocia quando uma proteína iniciadora se liga aos sítios de ligação adjacentes. Contrariamente aos cromossomos procarióticos, cada cromossomo eucariótico apresenta muitas origens de replicação para replicar rapidamente os genomas eucarióticos muito maiores. Aproximadamente 400 origens de replicação estão dispersas pelos 16 cromossomos de levedura e estima-se que existam milhares de forquilhas de replicação nos 23 cromossomos dos seres humanos. Portanto, em eucariotos, a replicação prossegue em ambos os sentidos a partir de múltiplos pontos de origem (Figura 7.23). As duplas-hélices que estão sendo produzidas em cada origem da replicação alongam e finalmente se unem umas às outras. Quando a replicação dos dois filamentos está completa, resultam duas **moléculas-filhas** de DNA idênticas.

> CONCEITO-CHAVE Onde e quando ocorre a replicação são cuidadosamente controlados pela montagem ordenada do replissomo em um local preciso, denominado origem. A replicação prossegue em ambos os sentidos a partir de uma única origem no cromossomo procariótico circular. A replicação prossegue em ambos os sentidos a partir de centenas de milhares de origens em cada um dos cromossomos eucarióticos lineares.

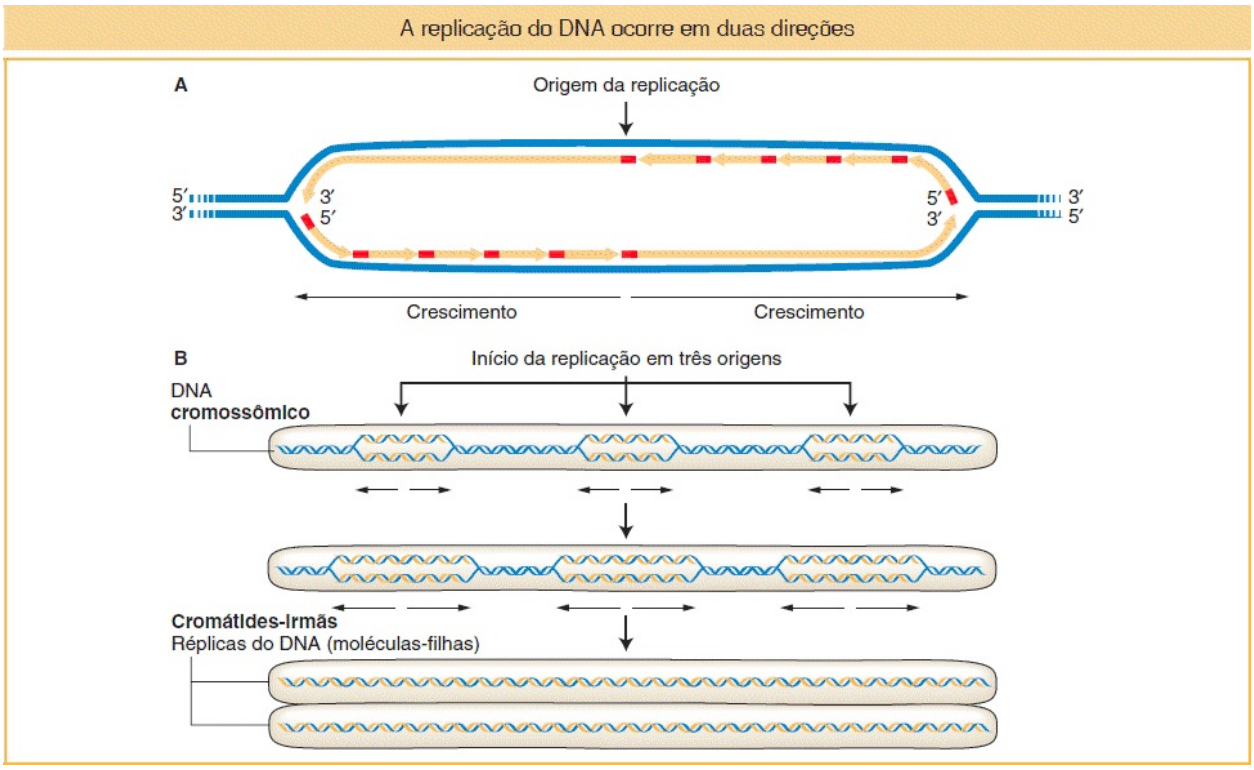


FIGURA 7.23 A replicação do DNA prossegue em ambos os sentidos a partir de uma origem de replicação. As *setas pretas* indicam o sentido do crescimento das moléculas-filhas de DNA. **A.** Começando na origem, as DNA polimerases se movimentam para o exterior em ambos os sentidos. As *setas amarelas longas* representam os filamentos contínuos e as *setas amarelas curtas* unidas representam os filamentos descontínuos. **B.** Como a replicação prossegue no nível cromossômico. Três origens de replicação estão demonstradas neste exemplo.

Replicação do DNA e ciclo celular de levedura

A síntese do DNA ocorre na fase S (síntese) do ciclo celular eucariótico (Figura 7.24). Como o início da síntese do DNA está limitado a esse estágio único? Em leveduras, o método de controle é ligar a montagem do replissomo ao ciclo celular. A Figura 7.25 demonstra o processo. Em leveduras, são necessárias três proteínas para iniciar a montagem do replissomo. O complexo de reconhecimento da origem (ORC) se liga primeiramente às sequências nas origens de leveduras, de modo muito semelhante como a proteína DnaA o faz em *E. coli*. A presença de ORC na origem atua para recrutar duas outras proteínas, Cdc6 e Cdt1. Ambas as proteínas mais o ORC em seguida recrutam a helicase, denominada complexo MCM, e os outros componentes do replissomo.

A replicação está ligada ao ciclo celular por meio da disponibilidade de Cdc6 e Cdt1. Em leveduras, essas proteínas são sintetizadas durante o final da mitose e o intervalo 1 (G1) e são destruídas por proteólise após o início da síntese. Desse modo, o replissomo pode ser montado apenas antes da fase S. Quando a replicação tem início, não pode haver formação de novos replissomos nas origens, tendo em vista que Cdc6 e Cdt1 são degradadas durante a fase S e deixam de estar disponíveis.

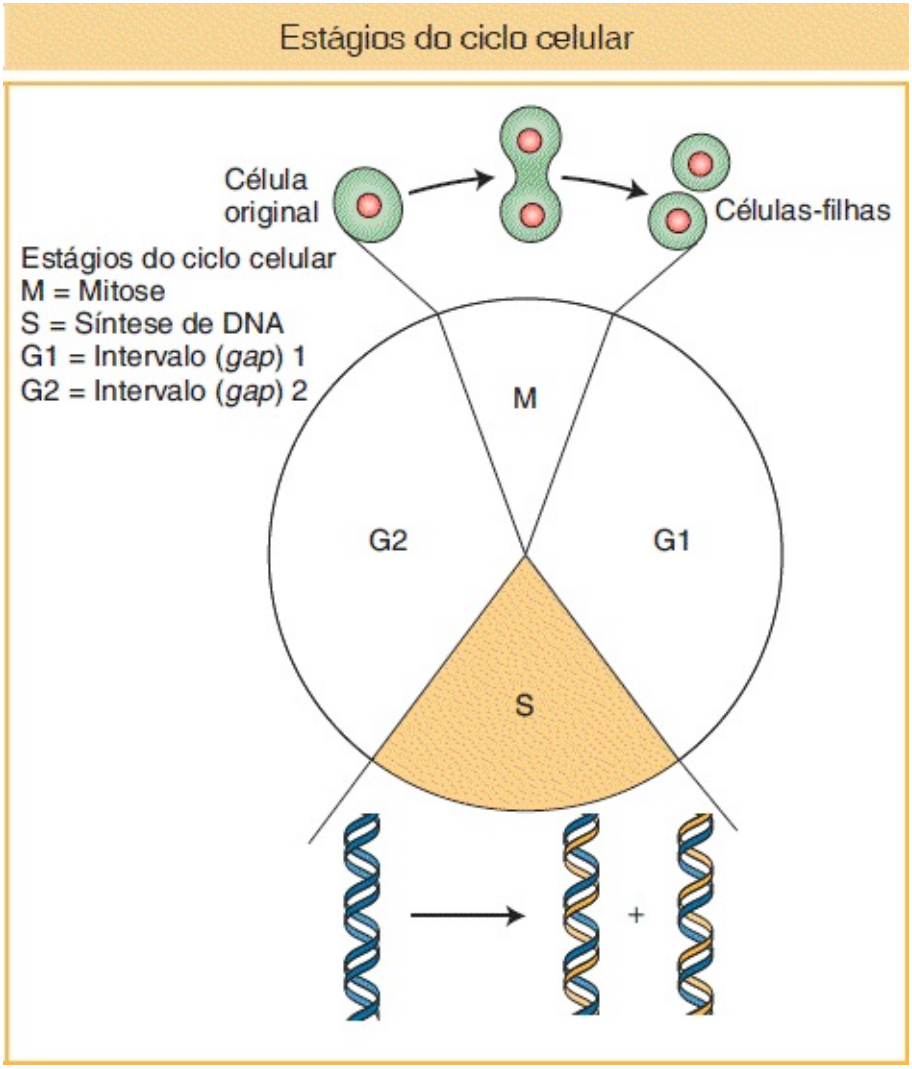


FIGURA 7.24 O DNA é replicado durante a fase S do ciclo celular.

Início da replicação eucariótica

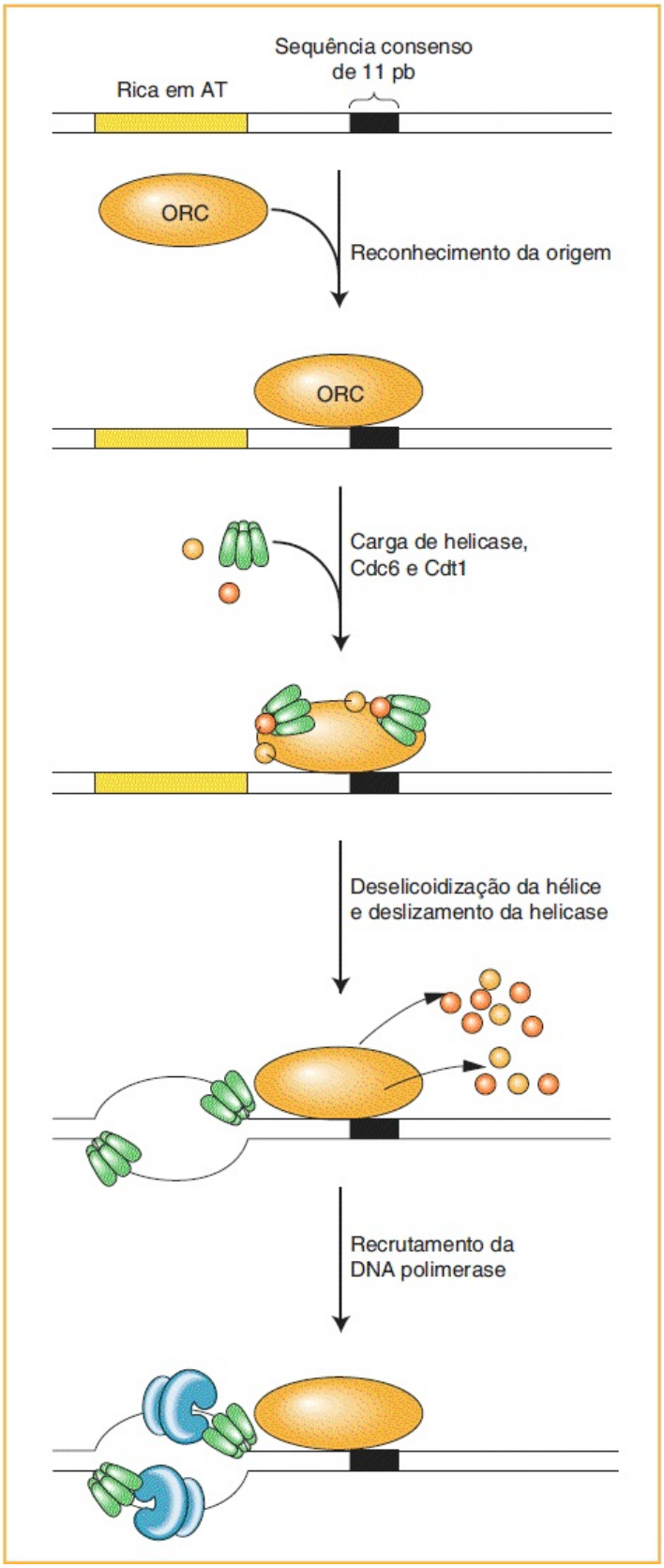


FIGURA 7.25 Este exemplo de levedura demonstra a iniciação da síntese de DNA em uma origem de replicação em um eucarioto. Assim como na iniciação da replicação procariótica (ver Figura 7.20), as proteínas do complexo de reconhecimento da origem (ORC) se ligam à origem, onde separam os dois filamentos da dupla-hélice e recrutam os componentes do replissomo nas duas forquilhas de replicação. A replicação está ligada ao ciclo celular por meio da disponibilidade de duas proteínas: Cdc6 e Cdt1.

Origens de replicação em eucariotos superiores

Conforme já declarado, a maior parte das aproximadamente 400 origens de replicação em leveduras é composta por sequências de DNA *motifs* semelhantes (100 a 200 pb de comprimento), que são reconhecidas pelas subunidades de ORC. Curiosamente, embora todos os eucariotos caracterizados apresentem proteínas ORC semelhantes, as origens de replicação em eucariotos superiores são muito mais longas, possivelmente de até dezenas de milhares ou centenas de milhares de nucleotídios. Significativamente, os eucariotos apresentam semelhança limitada de sequências. Portanto, embora o ORC de levedura reconheça se-quências de DNA específicas em cromossomos de levedura, o que os ORC correlatos de eucariotos superiores reconhecem não está claro nesse momento, mas a característica reconhecida provavelmente não é uma sequência de DNA específica. O que essa incerteza significa em termos práticos é que é muito mais difícil isolar origens de replicação de seres humanos e outros eucariotos superiores, tendo em vista que os cientistas não conseguem utilizar uma sequência de DNA isolada de uma origem humana, por exemplo, para realizar uma pesquisa computadorizada de toda a sequência do genoma humano para encontrar outras origens.

Se os ORC de eucariotos superiores são interagem com uma sequência específica dispersa pelos cromossomos, então como eles encontram as origens de replicação? Acredita-se que esses ORC interajam indiretamente com as origens por meio da associação com outros complexos proteicos que estão ligados aos cromossomos. Um referido mecanismo de reconhecimento pode ter evoluído de modo que os eucariotos superiores possam regular o momento da replicação do DNA durante a fase S (ver Capítulo 12 para mais a respeito da eucromatina e da heterocromatina). Há algum tempo sabe-se que as regiões cromossômicas ricas

em genes (a eucromatina) são replicadas inicialmente na fase S, enquanto as regiões pobres em genes, incluindo a heterocromatina condensada, são replicadas tardiamente na fase S. A replicação do DNA não poderia ser programada por região se os ORC estivessem ligados às sequências relacionadas dispersas pelos cromossomos. Em vez disso, os ORC, por exemplo, apresentam uma afinidade maior por origens na cromatina aberta e se ligar a essas origens primeiro e posteriormente se ligar à cromatina condensada apenas após as regiões ricas em genes terem sido replicadas.

> CONCEITO-CHAVE A origem da replicação de levedura, assim como a origem em procariotos, contém uma sequência de DNA conservada que é reconhecida pelo ORC e outras proteínas necessárias para montar o replissomo. Em contrapartida, as origens de eucariotos superiores têm sido difíceis de isolar e estudar, tendo em vista que são longas e complexas e não contêm uma sequência de DNA conservada.

7.7 Telômeros e telomerase | Término da replicação

A replicação da molécula de DNA linear em um cromossomo eucariótico prossegue em ambos os sentidos a partir de diversas origens de replicação, conforme demonstrado na Figura 7.23. Esse processo replica a maior parte do DNA cromossômico, mas existe um problema inerente na replicação das duas extremidades das moléculas de DNA lineares, as regiões denominadas **telômeros**. A síntese contínua no filamento *leading* pode prosseguir até a ponta do molde. Entretanto, a síntese do filamento descontínuo necessita de *primers* à frente do processo; assim, quando o último *primer* é removido, faltam sequências no final daquele filamento. Consequentemente, uma ponta unifilamentar permanece em uma das moléculas-filhas de DNA (Figura 7.26). Se o cromossomo-filho com essa molécula de DNA fosse replicado novamente, o filamento com sequências ausentes na extremidade se tornaria uma molécula bifilamentar encurtada após a replicação. A cada ciclo de replicação subsequente,

o telômero continuaria a encurtar, até que finalmente seriam perdidas informações codificantes essenciais.

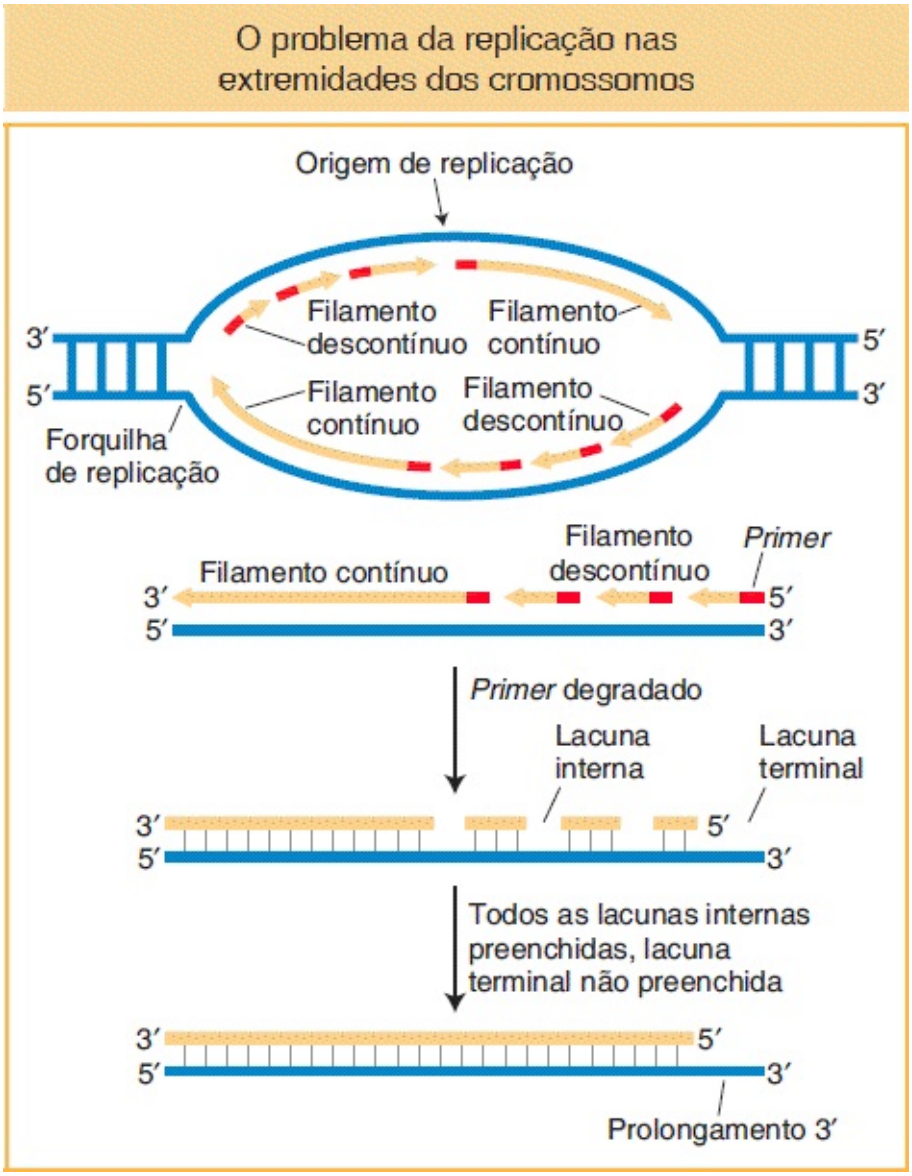


FIGURA 7.26 (*Parte superior*) A replicação de cada fragmento de Okazaki no filamento descontínuo tem início com a inserção de um *primer*. (*Parte inferior*) O destino do filamento inferior na bolha de transcrição. Quando o *primer* do último fragmento de Okazaki do filamento descontínuo é removido, não há como preencher a lacuna por meio da replicação convencional. Resultaria um cromossomo encurtado quando o cromossomo que contém a lacuna fosse replicado.

As células desenvolveram um sistema especializado para evitar essa perda. A solução envolve a adição de múltiplas cópias de uma sequência não codificadora

simples ao DNA nas pontas do cromossomo. Portanto, cada vez que um cromossomo é duplicado, ele fica mais curto e apenas essas sequências repetidas, que não contêm informações, são perdidas. As repetições perdidas são, em seguida, adicionadas novamente às extremidades do cromossomo.

A descoberta de que as extremidades dos cromossomos são compostas por sequências repetidas em *tandem* foi feita em 1978 por Elizabeth Blackburn e Joe Gall, que estavam estudando o DNA no macronúcleo incomum do ciliado unicelular *Tetrahymena*. Assim como outros ciliados, o *Tetrahymena* apresenta um micronúcleo convencional e um macronúcleo incomum, no qual os cromossomos são fragmentados em milhares de pedaços do tamanho de um gene com novas extremidades adicionadas a cada pedaço. Com tantas extremidades cromossômicas, *Tetrahymena* apresenta aproximadamente 40.000 telômeros e, como tal, era a escolha perfeita para determinar a composição do telômero. Blackburn e Gall foram capazes de isolar os fragmentos que continham os genes para RNA ribossômico (fragmentos denominados rDNA; ver Capítulo 9 para mais sobre os ribossomos) por meio da utilização da centrifugação em gradiente de CsCl, a técnica desenvolvida por Meselson e Stahl para isolar DNA de *E. coli* recém-replicado (ver anteriormente). As extremidades dos fragmentos de rDNA continham arranjos em *tandem* da sequência TTGGGG. Atualmente sabemos que virtualmente todos os eucariotos apresentam repetições curtas em *tandem* nas extremidades de seus cromossomos; entretanto, a sequência não é exatamente a mesma. Os cromossomos humanos, por exemplo, terminam em aproximadamente 10 a 15 kb de repetições em *tandem* da sequência TTAGGG.

A questão sobre como essas repetições de fato são adicionadas às extremidades dos cromossomos após cada rodada de replicação foi abordada por Elizabeth Blackburn e Carol Grieder. Elas formularam a hipótese de que uma enzima catalisava o processo. Novamente trabalhando com extratos do macronúcleo de *Tetrahymena*, elas identificaram uma enzima, que denominaram **telomerase**, que adiciona repetições curtas às extremidades 3′ das moléculas de DNA. Curiosamente, a proteína telomerase carrega uma pequena molécula de RNA, parte da qual atua como um molde para a síntese da unidade de repetição telomérica. Em todos os vertebrados, incluindo seres humanos, a sequência de

RNA 3′-AAUCCC-5′ atua como o molde para a unidade de repetição 5′-TTAGGG-3′ por meio de um mecanismo demonstrado na Figura 7.27. Brevemente, a RNA telomerase primeiro se liga ao prolongamento 3′ do DNA, que em seguida é estendido com a utilização dos dois componentes da telomerase: o pequeno RNA (como molde) e a proteína (como atividade de polimerase). Após a adição de alguns nucleotídios ao prolongamento 3′, a RNA telomerase se movimenta ao longo do DNA, de modo que a extremidade 3′ possa ser adicionalmente estendida por meio da sua atividade de polimerase. A extremidade 3′ continua a ser estendida pela movimentação repetida da RNA telomerase. A primase e a DNA polimerase em seguida utilizam o prolongamento 3′ muito longo como molde para preencher a extremidade do outro filamento de DNA. Trabalhando com Elizabeth Blackburn, um terceiro pesquisador, Jack Szostak, prosseguiu até demonstrar que também existem telômeros na levedura. Pela contribuição para a descoberta de como os telômeros protegem os cromossomos do encurtamento, Blackburn, Grieder e Szostak receberam o Prêmio Nobel em Medicina ou Fisiologia de 2009.

Alongamento do telômero

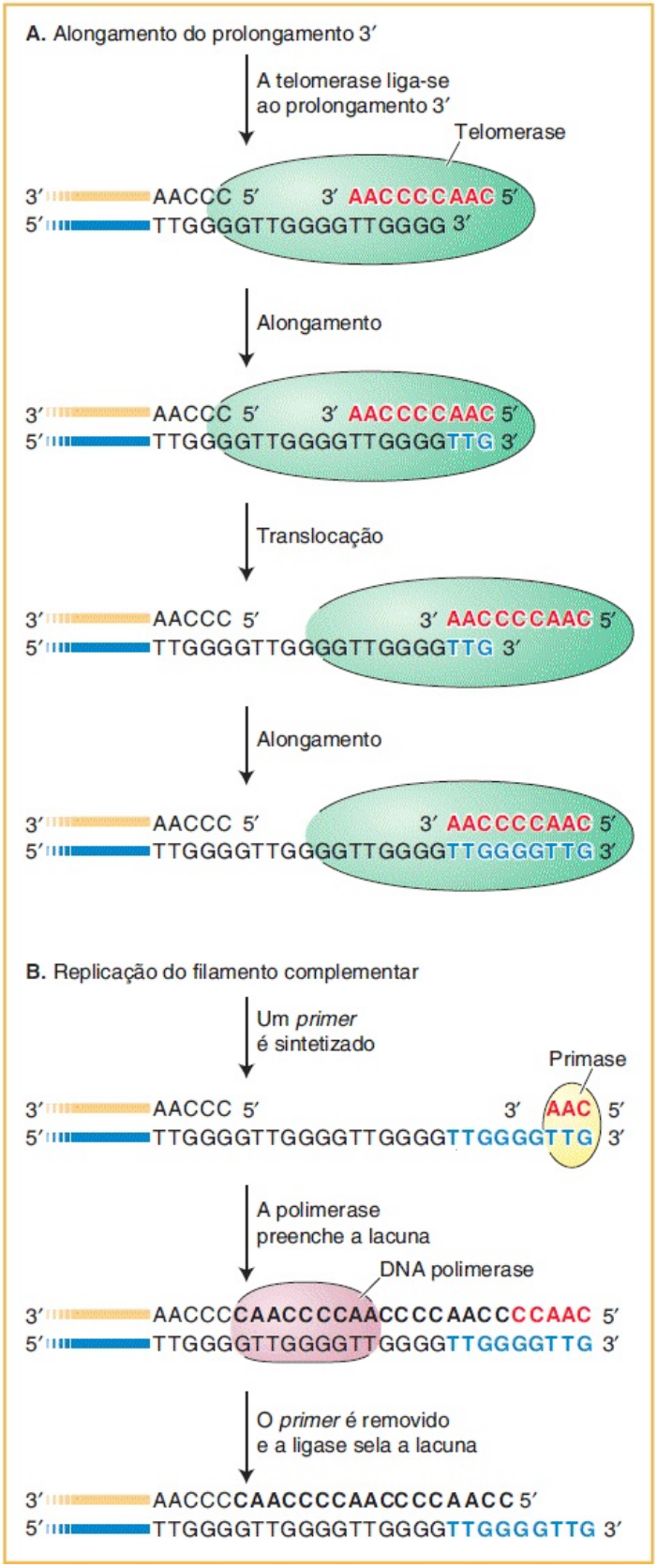


FIGURA 7.27 A telomerase carreia uma molécula de RNA curta (*letras vermelhas*), que atua como um molde para a adição de uma sequência de DNA complementar, que é adicionada ao prolongamento 3′ (*letras azuis*). Para adicionar outra repetição, a telomerase transloca-se para a extremidade da repetição que acabou de adicionar. O prolongamento 3′ estendido em seguida pode atuar como um molde para a replicação do DNA convencional.

Uma característica notável dessa reação é que o RNA serve como o molde para a síntese de DNA. Conforme você verificou no Capítulo 1 (e revisará no Capítulo 8), o DNA normalmente atua como molde para a síntese de RNA no processo denominado transcrição. É por esse motivo que se diz que a polimerase da telomerase apresenta atividade de transcriptase reversa. Revisaremos a transcriptase reversa nos Capítulos 10 e 15.

Além de prevenir a erosão do material genético após cada rodada de replicação, os telômeros preservam a integridade cromossômica por meio da associação com proteínas para formar capuzes protetores. Esses capuzes isolam o prolongamento unifilamentar 3′, que pode ter até 100 nucleotídios de comprimento (Figura 7.28). Sem esse capuz, as extremidades bifilamentares dos cromossomos seriam confundidas pela célula com quebras bifilamentares e seriam tratadas de acordo. Conforme você verá posteriormente, no Capítulo 16, quebras bifilamentares são possivelmente muito perigosas, tendo em vista que podem resultar em instabilidade cromossômica que pode levar ao câncer e a uma diversidade de fenótipos associados ao envelhecimento. Por esse motivo, quando uma quebra bifilamentar é detectada, a célula responde de vários modos, dependendo, em parte, do tipo celular e da extensão da lesão. Por exemplo, a quebra bifilamentar pode fundir-se com outra quebra ou a célula pode limitar o dano para o organismo interrompendo a divisão celular (denominada senescência) ou iniciando uma via de morte celular (denominada apoptose).

> CONCEITO-CHAVE Os telômeros são estruturas especializadas nas extremidades dos cromossomos que contêm repetições em *tandem* de uma sequência curta de DNA, que é adicionada à extremidade 3’ pela enzima telomerase. Os telômeros estabilizam os cromossomos ao prevenir a perda de

> informação genômica após cada rodada de replicação do DNA e por meio da associação com proteínas para formar um capuz que “esconde” as extremidades dos cromossomos do maquinário de reparo do DNA da célula.

Surpreendentemente, embora a maior parte das células germinativas apresente telomerase em abundância, as células somáticas produzem muito pouca ou nenhuma telomerase. Por esse motivo, os cromossomos de células somáticas proliferativas se tornam progressivamente mais curtos a cada divisão celular, até que a célula interrompe todas as divisões e entra em uma fase de senescência. Essa observação levou muitos investigadores a suspeitarem que havia uma ligação entre o encurtamento dos telômeros e o envelhecimento. Geneticistas que estudam doenças humanas que levam ao fenótipo de envelhecimento precoce recentemente revelaram evidências que amparam essa conexão. Pessoas com síndrome de Werner apresentam manifestação precoce de muitos eventos relacionados com a idade, incluindo enrugamento da pele, catarata, osteoporose, branqueamento dos cabelos e doença cardiovascular (Figura 7.29). Estudos genéticos e bioquímicos observaram que pessoas afetadas apresentam telômeros mais curtos do que aqueles de pessoas normais em virtude de uma mutação em um gene denominado WRN, que codifica uma proteína (uma helicase) que está associada a proteínas que formam a estrutura de revestimento do telômero (TRF2, ver Figura 7.28). Formula-se a hipótese de que essa mutação rompe o telômero normal, resultando em instabilidade cromossômica e no fenótipo de envelhecimento precoce. Pacientes com outra síndrome de envelhecimento precoce denominada disqueratose congênita também apresentam telômeros mais curtos do que aqueles de pessoas saudáveis da mesma idade e também apresentam mutações em genes necessários para a atividade da telomerase.

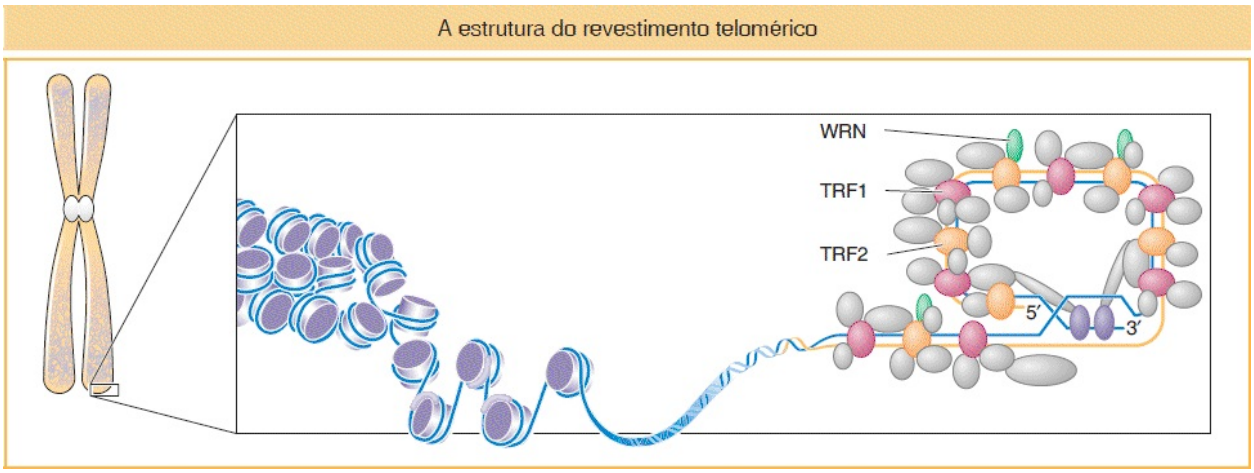


FIGURA 7.28 Um “capuz” protege o telômero na extremidade de um cromossomo. O prolongamento 3′ é “escondido” quando desloca um filamento de DNA em uma região na qual as repetições teloméricas são bifilamentares. As proteínas TRF1 e TRF2 se ligam às repetições teloméricas e outras proteínas, incluindo WRN, ligam-se a TRF1 e TRF2, formando, assim, o capuz protetor do telômero.

Os geneticistas também estão muito interessados em conexões entre telômeros e câncer. Contrariamente às células somáticas normais, a maior parte das células cancerosas apresenta atividade de telomerase. A capacidade de manter telômeros funcionais pode ser um motivo pelo qual as células cancerosas, mas não as células normais, conseguem crescer em culturas celulares durante décadas e são consideradas imortais. Como tal, muitas empresas farmacêuticas estão buscando capitalizar essa diferença entre as células cancerosas e normais com o desenvolvimento de fármacos que têm por alvo seletivo as células cancerosas por meio da inibição da atividade da telomerase.

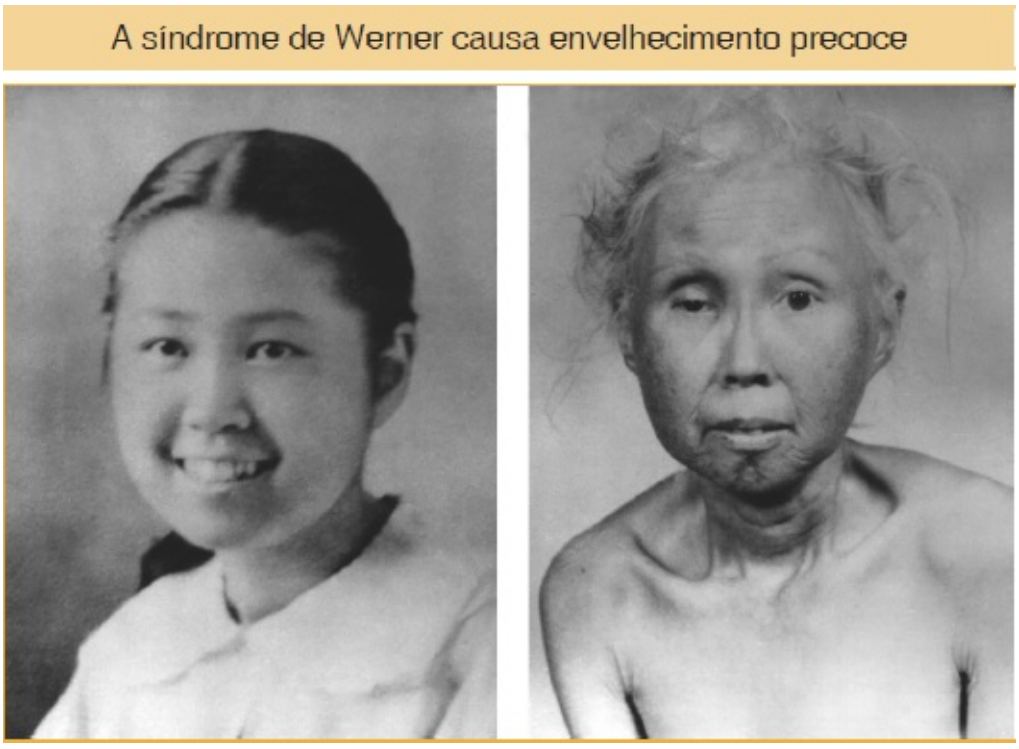


FIGURA 7.29 Uma mulher com síndrome de Werner aos 15 e aos 48 anos de idade. (International Registry of Werner Syndrome, www.wernersyndrome.org.)

RESUMO

Trabalhos experimentais sobre a natureza molecular do material hereditário demonstraram conclusivamente que o DNA (e não proteínas, lipídios ou carboidratos) é de fato o material genético. Com a utilização de dados obtidos por outros pesquisadores, Watson e Crick deduziram um modelo de dupla-hélice com dois filamentos de DNA, enrolados um ao redor do outro, de modo antiparalelo. A ligação dos dois filamentos tem por base o pareamento da adenina (A) com a timina (T) e da guanina (G) com a citosina (C). O primeiro par é unido por duas ligações de hidrogênio; o segundo, por três.

O modelo de Watson-Crick demonstra como o DNA pode ser replicado de modo ordenado — requisito indispensável para o material genético. A replicação é realizada de modo semiconservativo tanto em procariotos quanto em eucariotos. Uma dupla-hélice é replicada para formar duas hélices idênticas, cada uma com seus nucleotídios na ordem linear idêntica; cada uma das duas novas duplas-

hélices é composta por um filamento antigo e um filamento recém-polimerizado de DNA.

A dupla-hélice de DNA é desenrolada em uma forquilha de replicação e os dois filamentos únicos assim produzidos atuam como moldes para a polimerização de nucleotídios livres. Os nucleotídios são polimerizados pela enzima DNA polimerase, que adiciona novos nucleotídios apenas à extremidade 3′ de uma cadeia de DNA em crescimento. Tendo em vista que a adição é apenas nas extremidades 3′, a polimerização em um molde é contínua, produzindo o filamento contínuo (*leading*) e, no outro, é descontínua em trechos curtos (fragmentos de Okazaki), produzindo o filamento descontínuo (*lagging*). A síntese do filamento contínuo e de cada fragmento de Okazaki é iniciada por um *primer* de RNA curto (sintetizado pela primase) que fornece uma extremidade 3′ para a adição de desoxirribonucleotídios.

Os eventos múltiplos que têm de ocorrer com precisão e rapidez na forquilha de replicação são realizados por uma máquina biológica denominada replissomo. Esse complexo proteico inclui duas unidades de DNA polimerase, uma para atuar no filamento contínuo e uma para atuar no filamento descontínuo. Desse modo, a síntese é mais demorada e a união dos fragmentos de Okazaki em um filamento contínuo pode ser temporalmente coordenada com a síntese menos complicada do filamento contínuo. Onde e quando a replicação ocorre é cuidadosamente controlado pela montagem ordenada do replissomo em determinados locais nos cromossomos, denominados origens. Os genomas eucarióticos podem apresentar dezenas de milhares de origens. A montagem dos replissomos nessas origens ocorre apenas em uma fase específica no ciclo celular.

As extremidades dos cromossomos lineares (telômeros) apresentam um problema para o sistema de replicação, tendo em vista que sempre há um trecho curto em um filamento que não pode ser iniciado. A enzima telomerase adiciona várias sequências curtas e repetitivas para manter o comprimento. A telomerase carrega um RNA curto que atua como o molde para a síntese das repetições teloméricas. Essas repetições teloméricas não codificadoras associam-se a proteínas para formar um revestimento telomérico. Os telômeros encurtam com a idade nas células somáticas, tendo em vista que a telomerase não é produzida

naquelas células. Indivíduos que apresentam telômeros defeituosos sofrem envelhecimento precoce.

TERMOS-CHAVE

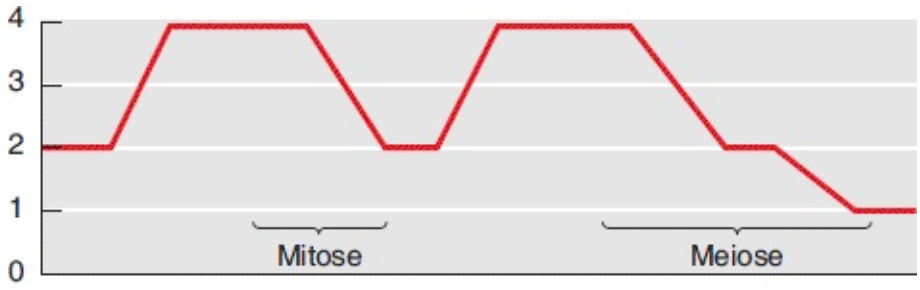
antiparalela
base
base complementar
ceto
código genético
desoxirribose
DNA ligase
dupla-hélice
enol
enzima distributiva
enzima processiva
filamento contínuo (líder, *leading*)
filamento descontínuo (lento, *lagging*)
forquilha de replicação
fosfato
fragmento de Okazaki
helicase
holoenzima polimerase III (pol III)
imino
modelo ou molde
molécula-filha
nucleotídio
origem
pirimidina
primase
primer
primossomo

proteína de ligação a filamento simples (SSB)
purina
replicação conservativa
replicação semiconservativa
replissomo
sulco maior
sulco menor
tautomerização
telomerase
telômero
topoisomerase

PROBLEMAS RESOLVIDOS

Problema resolvido 1. A mitose e a meiose foram apresentadas no Capítulo 2. Considerando o que foi abrangido neste capítulo a respeito da replicação do DNA, desenhe um gráfico demonstrando o conteúdo de DNA contra o tempo em uma célula que sofre mitose e em seguida meiose. Presuma uma célula diploide.

Solução



Problema resolvido 2. Se o conteúdo de GC de uma molécula de DNA é de 56%, quais são as porcentagens das quatro bases (A, T, G e C) nessa molécula?

Solução

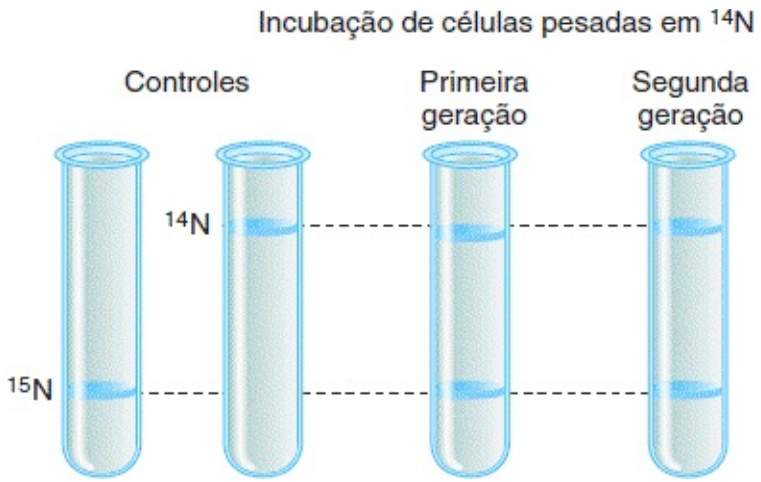
Se o conteúdo de GC for de 56%, então, tendo em vista que G = C, o conteúdo de G é de 28% e o conteúdo de C é de 28%. O conteúdo de AT = 100 — 56 = 44%.

Tendo em vista que A = T, o conteúdo de A é de 22% e o conteúdo de T é de 22%.

Problema resolvido 3. Descreva o padrão esperado de bandas em um gradiente de CsCl para a replicação *conservativa* no experimento de Meselson-Stahl. Desenhe um diagrama.

Solução

Consulte a Figura 7.13 para uma explicação adicional. Na replicação conservativa, se bactérias são cultivadas na presença de ^{15}N e em seguida transferidas para ^{14}N, uma molécula de DNA será toda ^{15}N após a primeira geração e a outra molécula será toda ^{14}N, resultando em uma banda pesada e uma banda leve no gradiente. Após a segunda geração, o DNA ^{15}N produzirá uma molécula toda com ^{15}N e uma molécula toda com ^{14}N, enquanto o DNA ^{14}N produzirá apenas DNA^{14}N. Portanto, apenas DNA todo ^{14}N ou todo ^{15}N é gerado, novamente produzindo uma banda leve e uma banda pesada:



PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Tabela 7.1, por que não há dados em relação às primeiras quatro fontes teciduais? Em relação às três últimas entradas, qual é a explicação mais

provável para as discretas diferenças na composição do DNA humano das três fontes teciduais?

2. Na Figura 7.7, você reconhece alguns dos componentes utilizados para a construção do modelo do DNA de Watson e Crick? Onde você os viu anteriormente?

3. Fazendo referência à Figura 7.20, responda às questões a seguir:
a. O que a enzima DNA polimerase I está fazendo?
b. Quais outras proteínas são necessárias para que a DNA polimerase III à esquerda continue sintetizando DNA?
c. Quais outras proteínas são necessárias para que a DNA polimerase III à direita continue sintetizando DNA?

4. O que há de diferente a respeito da reação catalisada pela helicase verde na Figura 7.20 e a girase amarela na Figura 7.21?

5. Na Figura 7.23 A, rotule todos os filamentos líder (*leading*) e lentos (*lagging*).

PROBLEMAS BÁSICOS

6. Descreva os tipos de ligações químicas na dupla-hélice do DNA.

7. Explique qual é o significado dos termos *replicação conservativa* e *semiconservativa*.

8. Qual é o significado de *primer* e por que os *primers* são necessários para a replicação do DNA?

9. O que são helicases e topoisomerases?

10. Por que a síntese do DNA é contínua em um filamento e descontínua no filamento oposto?

11. Se os quatro desoxinucleotídios demonstrassem pareamento inespecífico de bases (A com C, A com G, T com G, e assim por diante), as informações únicas contidas em um gene seriam mantidas a cada ciclo de replicação? Explique.

12. Se não houvesse helicases durante a replicação, o que ocorreria com o processo de replicação?

13. O que aconteceria se, durante a replicação, as topoisomerases fossem incapazes de reconectar os fragmentos de DNA de cada filamento após a a deselicoidização (relaxamento) da molécula de DNA?

14. Qual das que seguem *não* é uma propriedade-chave do material hereditário?

a. Ele tem de ser capaz de ser copiado com precisão.

b. Ele obrigatoriamente codifica as informações necessárias para formar proteínas e estruturas complexas.

c. Ele sofre mutação ocasionalmente.

d. Ele precisa ser capaz de se adaptar a cada um dos tecidos do corpo.

15. É essencial que os *primers* de RNA nas extremidades dos fragmentos de Okazaki sejam removidos e substituídos por DNA, porque, de outro modo, qual dos eventos a seguir resultaria?

a. O RNA interferiria com a função da topoisomerase.

b. O RNA apresentaria maior probabilidade de conter erros, tendo em vista que a primase não apresenta uma função de revisão.

c. O grampo β do dímero da DNA pol II liberaria o DNA e a replicação seria interrompida.

d. Os *primers* de RNA provavelmente formariam ligações de hidrogênio entre si, originando estruturas complexas que podem interferir com a formação adequada da hélice de DNA.

16. As polimerases normalmente adicionam apenas aproximadamente 10 nucleotídios a um filamento de DNA antes de se dissociarem. Entretanto, durante a replicação, a DNA pol III consegue adicionar dezenas de milhares de nucleotídios a uma forquilha em movimento. Como essa adição é realizada?

17. Em cada origem de replicação, a síntese do DNA prossegue bidirecionalmente a partir de duas forquilhas de replicação. Qual dos que seguem ocorreria se um mutante eucariótico surgisse e apresentasse apenas

uma forquilha funcional por bolha de replicação? (Ver diagrama.)

a. Nenhuma alteração na replicação.
b. A replicação ocorreria apenas em metade do cromossomo.
c. A replicação seria completa apenas no filamento contínuo.
d. A replicação demoraria o dobro do tempo.

18. Em uma célula diploide na qual $2n = 14$, quantos telômeros existem em cada uma das fases a seguir do ciclo celular?
(a) G1
(b) G2
(c) prófase da mitose
(d) telófase da mitose

19. Se a timina compõe 15% das bases em uma molécula de DNA específica, qual porcentagem das bases é de citosina?

20. Se o conteúdo de GC de uma molécula de DNA é de 48%, quais são as porcentagens das quatro bases (A, T, G e C) nessa molécula?

21. Bactérias denominadas extremófilas são capazes de crescer em fontes termais, tais como o gêiser Old Faithful, no Parque Nacional Yellowstone, em Wyoming. Você acredita que o DNA das extremófilas apresenta conteúdo mais alto de pares de bases GC ou AT? Justifique a sua resposta.

22. Presuma que um determinado cromossomo bacteriano apresenta uma origem de replicação. Sob algumas condições de divisão celular rápida, a replicação poderia ter início a partir da origem antes de o ciclo de replicação precedente ser concluído. Quantas forquilhas de replicação estariam presentes sob essas condições?

23. Uma molécula de composição:

5′-AAAAAAAAAAA-3′
3′-TTTTTTTTTTTTT-5′

é replicada em uma solução que contém GTP, CTP e TTP não marcados (não radioativo), mais nucleosídios adenosina trifosfato com todos seus átomos de fósforo na forma do isótopo radioativo ^{32}P. Ambas as moléculas-filhas serão radioativas? Explique. Em seguida repita a questão em relação à molécula:

5′-ATATATATATATAT-3′
3′-TATATATATATATA-5′

24. O experimento de Meselson e Stahl teria funcionado se houvessem sido utilizadas células eucarióticas diploides?

25. Considere o segmento de DNA a seguir, que é parte de uma molécula muito mais longa que constitui um cromossomo:

5′....ATTCGTACGATCGACTGACTGACAGTC....3′
3′....TAAGCATGCTAGCTGACTGACTGTCAG....5′

Se a DNA polimerase iniciar a replicação desse segmento a partir da direita:

a. Qual será o molde para o filamento contínuo?

b. Desenhe a molécula quando a DNA polimerase está na metade desse segmento.

c. Desenhe as duas moléculas-filhas completas.

d. O seu diagrama na parte *b* é compatível com replicação bidirecional a partir de uma origem única, o modo de replicação habitual?

26. As DNA polimerases são posicionadas ao longo do segmento de DNA a seguir (que é parte de uma molécula muito maior) e se movimentam da direita para a esquerda. Se presumirmos que um fragmento de Okazaki é produzido a partir desse segmento, qual será a sequência do fragmento? Rotule suas extremidades 5′ e 3′.

5′....CCTTAAGACTAACATCTTACTGGGATC....3′
3′....GGAATTCTGATTGATGAATGACCCTAG....5′

27. Permite-se que cromossomos de *E. coli* nos quais cada um dos átomos de nitrogênio está marcado (ou seja, cada átomo de nitrogênio é o isótopo pesado ^{15}N em vez do isótopo normal ^{14}N) repliquem em um ambiente no qual todo nitrogênio é ^{14}N. Com a utilização de uma linha sólida para representar uma cadeia pesada de polinucleotídios e uma linha tracejada para uma cadeia leve, esboce cada uma das descrições a seguir:

a. O cromossomo parental pesado e os produtos da primeira replicação após a transferência para um meio ^{14}N, presumindo que o cromossomo é uma dupla-hélice de DNA e que a replicação é semiconservativa.

b. Repita a parte *a*, mas agora presuma que a replicação é conservativa.

c. Se os cromossomos-filhos da primeira divisão em ^{14}N forem centrifugados em um gradiente de densidade de cloreto de sódio e uma banda única for obtida, qual das possibilidades nas partes *a* e *b* pode ser descartada? Reconsidere o experimento de Meselson e Stahl. O que ele *comprova*?

PROBLEMAS DESAFIADORES

28. Se uma mutação que inativou a telomerase ocorreu em uma célula (atividade da telomerase na célula = zero), qual você espera que seja o desfecho?

29. No planeta Rama, o DNA é de seis tipos de nucleotídios: A, B, C, D, E e F. Os tipos A e B são denominados *marzines*, C e D são *orsines*, E e F são *pirines*. As regras a seguir são válidas em todos os DNA de Rama:

Total de marzines = Total de orsines = Total de pirines

$$A = C = E$$
$$B = D = F$$

a. Prepare um modelo em relação à estrutura do DNA de Rama.

b. Em Rama, a mitose produz três células-filhas. Tendo esse fato em mente, proponha um padrão de replicação para o seu modelo de DNA.

c. Considere o processo da meiose em Rama. Quais comentários ou conclusões você pode sugerir?

30. Se você extrair o DNA do colifago ϕ X174, você observará que a sua composição é de 25% A, 33% T, 24% G, e 18% C. Essa composição é compatível com as regras de Chargaff? Como você interpretaria esse resultado? Como esse fago poderia replicar o seu DNA?

31. No Capítulo 5, você viu que as bactérias transferem DNA de um membro de sua espécie para outro em um processo denominado conjugação. Recentemente foi demonstrado que a transferência do DNA de uma célula bacteriana para outra não está limitada aos membros da mesma espécie. Um microbiologista que estuda a bactéria *Diplococcus pneumoniae* formula a hipótese de que uma região do seu cromossomo foi de fato transferida do *Mycobacterium tuberculosis*. Com base nos dados apresentados na Tabela 7.1, que característica distintiva do DNA transferido forneceria amparo para essa hipótese?

32. Tendo em vista o que você sabe a respeito da estrutura e da função da telomerase, forneça um modelo plausível para explicar como uma espécie poderia existir com uma combinação de duas repetições diferentes (p. ex., TTAGGG e TTGTGG) em cada um de seus telômeros.

33. As bactérias necessitam de telomerase? Explique por que sim ou por que não.

34. Watson e Crick utilizaram uma abordagem denominada construção de modelo para deduzir a estrutura da dupla-hélice do DNA. Como ela difere da abordagem experimental mais convencional que é realizada em um laboratório de pesquisas? Nesse sentido, por que o experimento de Meselson e Stahl foi considerado de tal importância crítica?

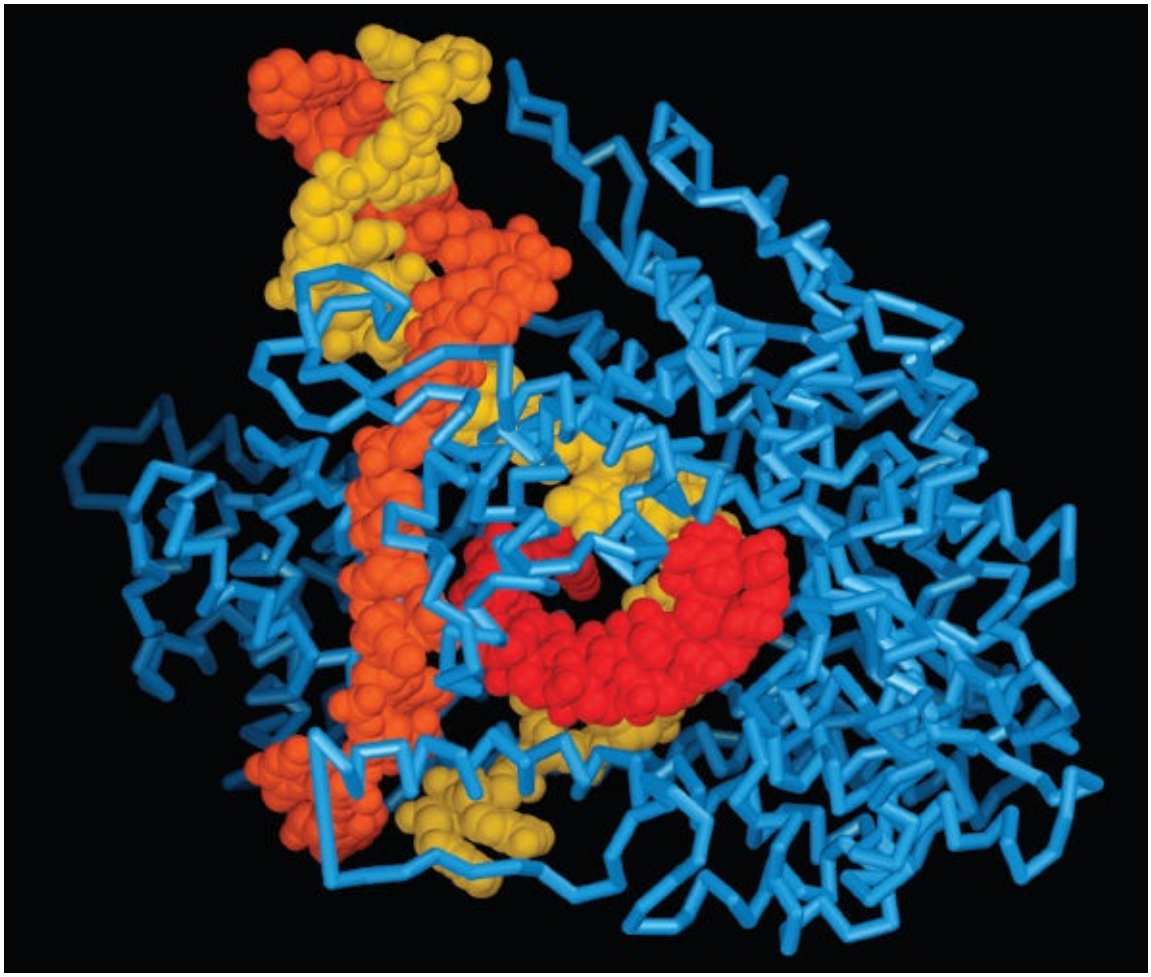
[1]J. Watson e F. Crick, *Nature* 171: 737, 1953.

[2]J. Watson e F. Crick, *Nature* 171:737, 1953.

CAPÍTULO

8

RNA | Transcrição e Processamento



A RNA polimerase em ação. Uma RNA polimerase (*azul*) muito pequena, produzida pelo bacteriófago T7, transcreve o DNA em um filamento de RNA (*vermelho*). A enzima separa a dupla-hélice do DNA (*amarelo*, *laranja*), expondo o filamento-molde para ser copiado em RNA (David S. Goodsell, Scripps Research Institute.)

TÓPICOS

8.1 RNA
8.2 Transcrição
8.3 Transcrição em eucariotos
8.4 Remoção de íntrons e recomposição de éxons
8.5 Pequenos RNA funcionais que regulam e protegem o genoma eucariótico

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Descrever como a estrutura do RNA difere daquela do DNA
- Distinguir as diferentes classes de RNA em uma célula
- Explicar a função dos promotores e as características necessárias para iniciar a transcrição
- Diagramar as etapas no processamento do RNA a partir de sua transcrição até o seu transporte para fora do núcleo
- Avaliar o motivo de a descoberta dos íntrons autorremovíveis ser considerada tão importante
- Descrever os diferentes tipos de RNA não codificadores (ncRNA).

Com a utilização do conhecimento recentemente adquirido sobre as sequências de DNA de genomas inteiros, cientistas têm sido capazes de determinar o número aproximado de genes em diversos organismos, tanto simples quanto complexos. Primeiramente não havia surpresas: a bactéria *Escherichia coli* apresenta aproximadamente 4.400 genes, a levedura *Saccharomyces cerevisiae*, um eucarioto unicelular, apresenta aproximadamente 6.300 genes, e a mosca-das-frutas multicelular *Drosophila melanogaster* apresenta aproximadamente 13.600 genes. Cientistas presumiram que organismos mais complexos necessitariam de mais genes e, assim, as estimativas iniciais eram de que o nosso genoma apresentaria 100.000 genes. Em uma conferência que focou em pesquisas genômicas, no ano 2000, cientistas iniciaram um grupo de

apostas informais denominado GeneSweep, que seria vencido por quem houvesse previsto de modo mais próximo o número real de genes no genoma humano. As apostas variaram de aproximadamente 26.000 a aproximadamente 150.000 genes.

Com a liberação do primeiro rascunho de sequência, foi anunciado um vencedor. Surpreendentemente, o vencedor era o apostador com a mais baixa estimativa, 25.947 genes. Como o *Homo sapiens*, com seu cérebro complexo e seu sistema imune sofisticado, apresenta apenas o dobro de genes dos nematódeos e aproximadamente o mesmo número de genes da primeira planta a ter seu genoma sequenciado, a erva daninha mostarda *Arabidopsis thaliana*? Parte da resposta a essa questão está relacionada com uma descoberta extraordinária feita no fim da década de 1970. Naquela ocasião, observou-se que as proteínas de muitos eucariotos são codificadas no DNA não como trechos contínuos (como ocorre em bactérias e leveduras), mas em segmentos. Portanto, os genes de eucariotos superiores normalmente são compostos por segmentos denominados **éxons** (que se referem às regiões *ex*pressas), que codificam partes das proteínas, e segmentos denominados **íntrons** (que se referem às regiões *in*tercalares), que separam os éxons. Como você aprenderá neste capítulo, uma cópia de RNA que contém tanto éxons quanto íntrons é sintetizada a partir de um gene. Uma máquina biológica (denominada **spliceossomo**) remove os íntrons e une os éxons (em um processo denominado **recomposição ou *splicing* do RNA**), para produzir um RNA maduro que contém a informação contínua necessária para sintetizar uma proteína.

Como os éxons e os íntrons estão relacionados com uma baixa contagem de genes humanos? Por enquanto, é suficiente dizer que o RNA transcrito a partir de um gene pode ser recomposto de modos alternativos. Embora apresentemos apenas aproximadamente 21.000 genes, esses genes codificam mais de 100.000 proteínas, graças ao processo de *splicing* alternativo do RNA.

Ainda mais surpreendente é o achado de que apenas uma pequena fração do genoma de fato codifica proteínas (pouco mais de 2% na maior parte dos organismos multicelulares complexos). O conteúdo dos genomas será o assunto de capítulos futuros. Por enquanto, é importante observar que apesar de apresentar uma pequena proporção de DNA codificador, a maior parte do genoma ainda codifica RNA. A história desse adequadamente denominado **RNA não**

codificador de proteína (ncRNA) é um trabalho em andamento. Aquela história será introduzida neste capítulo e será desenvolvida nos capítulos posteriores.

Neste capítulo, observamos as primeiras etapas na transferência de informações dos genes para os produtos gênicos. Na sequência do DNA do genoma de qualquer organismo está codificada a informação que especifica cada um dos produtos gênicos que o organismo pode produzir. Essas sequências de DNA também contêm informações que especificam quando, onde e quanto produto é produzido. Para utilizar a informação, uma cópia de RNA do gene deve ser sintetizada em um processo denominado *transcrição*.

A transferência da informação do gene para o produto gênico ocorre em diversas etapas. A primeira etapa, que é o foco deste capítulo, é copiar (*transcrever*) a informação em um filamento de RNA com a utilização do DNA como um molde. Em procariotos, a informação no RNA codificador de proteína é quase imediatamente convertida em uma cadeia de aminoácidos (proteína) por meio de um processo denominado *tradução*. Essa segunda etapa é o foco do Capítulo 9. Em eucariotos, a transcrição e a tradução são separadas espacialmente: a transcrição ocorre no núcleo e a tradução no citoplasma. Entretanto, antes que os RNA estejam prontos para ser transportados para o citoplasma para tradução ou outros usos, eles são submetidos a um processamento extensivo, incluindo a remoção de íntrons e a adição de um revestimento (*cap*) especial em 5′ e uma cauda de nucleotídios adenina em 3′. Um tipo de RNA totalmente processado, denominado *RNA mensageiro* (*mRNA*), é o intermediário na síntese das proteínas. Além disso, em procariotos e eucariotos, existem outros tipos de RNA que nunca são traduzidos. Esses ncRNA realizam muitos papéis essenciais.

A função do DNA e do RNA durante a transcrição tem por base dois princípios:

1. A complementaridade de bases é responsável pela determinação da sequência do transcrito de RNA na transcrição. Por meio do pareamento das bases complementares, a informação codificada no DNA passa para o RNA, e os complexos proteicos associados aos ncRNA são guiados até regiões

específicas no RNA para regular a sua expressão.

2. Determinadas proteínas reconhecem sequências de bases específicas no DNA e no RNA. Essas proteínas de ligação a ácidos nucleicos se ligam a essas sequências e atuam nelas.

Veremos esses princípios em atuação durante as discussões detalhadas sobre a transcrição e a tradução que se seguem neste capítulo e nos capítulos posteriores.

> CONCEITO-CHAVE As interações do DNA e do RNA ocorrem por meio do pareamento de bases complementares e da ligação de diversas proteínas a sítios específicos no DNA ou no RNA.

8.1 RNA

Investigadores iniciais tinham bons motivos para acreditar que a informação não é transferida diretamente do DNA para a proteína. Em uma célula eucariótica, o DNA é observado no núcleo, enquanto a proteína é sintetizada no citoplasma. É necessário um intermediário.

Experimentos iniciais sugerem um RNA intermediário

Em 1957, Elliot Volkin e Lawrence Astrachan fizeram uma observação significativa. Eles observaram que uma das alterações moleculares mais surpreendentes, que ocorre quando a *E. coli* é infectada pelo fago T2, é uma salva de síntese de RNA. Além disso, esse RNA induzido por fago "se renova" rapidamente; ou seja, seu tempo de vida é breve, da ordem de minutos. Seus rápidos aparecimento e desaparecimento sugeriram que o RNA pode desempenhar algum papel na expressão do genoma de T2 necessário para produzir mais partículas virais.

Volkin e Astrachan demonstraram a rápida rotatividade do RNA por meio da utilização de um protocolo denominado **experimento de pulso-caça**. Para conduzir um experimento de pulso-caça, as bactérias infectadas primeiramente

são alimentadas (pulso) com uracila radioativa (uma molécula necessária para a síntese de RNA, mas não de DNA). Qualquer RNA sintetizado nas bactérias a partir de então é "marcado" com uracila radioativa prontamente detectável. Após um curto período de incubação, a uracila radioativa é lavada e substituída (caça) pela uracila não radioativa. Esse procedimento "caça" a marcação para fora do RNA, tendo em vista que, na medida em que o RNA marcado com o pulso é degradado, apenas os precursores não marcados estão disponíveis para sintetizar novas moléculas de RNA (os nucleotídios marcados são "diluídos" pelo enorme excesso de uracila não marcada adicionado na caça). O RNA recuperado brevemente após o pulso é marcado, mas o RNA recuperado algum tempo depois não é marcado, indicando que o RNA apresenta um período de vida muito curto.

Um experimento semelhante pode ser realizado com células eucarióticas. As células primeiramente são pulsadas com uracila radioativa e, após um curto período, são transferidas para um meio com uracila não marcada. Em amostras coletadas após o pulso, a maior parte da marcação está no núcleo. Em amostras coletadas após a caça, o RNA marcado também é encontrado no citoplasma (Figura 8.1). Aparentemente, em eucariotos, o RNA é sintetizado no núcleo e em seguida se movimenta para o citoplasma, onde as proteínas são sintetizadas. Portanto, o RNA é um bom candidato para ser um intermediário na transferência de informação entre o DNA e a proteína.

Propriedades do RNA

Consideraremos as características gerais do RNA. Embora ambos o RNA e o DNA sejam ácidos nucleicos, o RNA difere do DNA em diversos aspectos importantes:

1. O RNA apresenta o açúcar **ribose** em seus nucleotídios, em vez da desoxirribose observada no DNA. Conforme o nome sugere, os dois açúcares diferem na presença ou na ausência de apenas um átomo de oxigênio. O açúcar do RNA contém um grupo hidroxila (OH) ligado ao átomo de carbono 2′, enquanto o açúcar do DNA apresenta apenas um átomo de hidrogênio ligado ao átomo de carbono 2′.

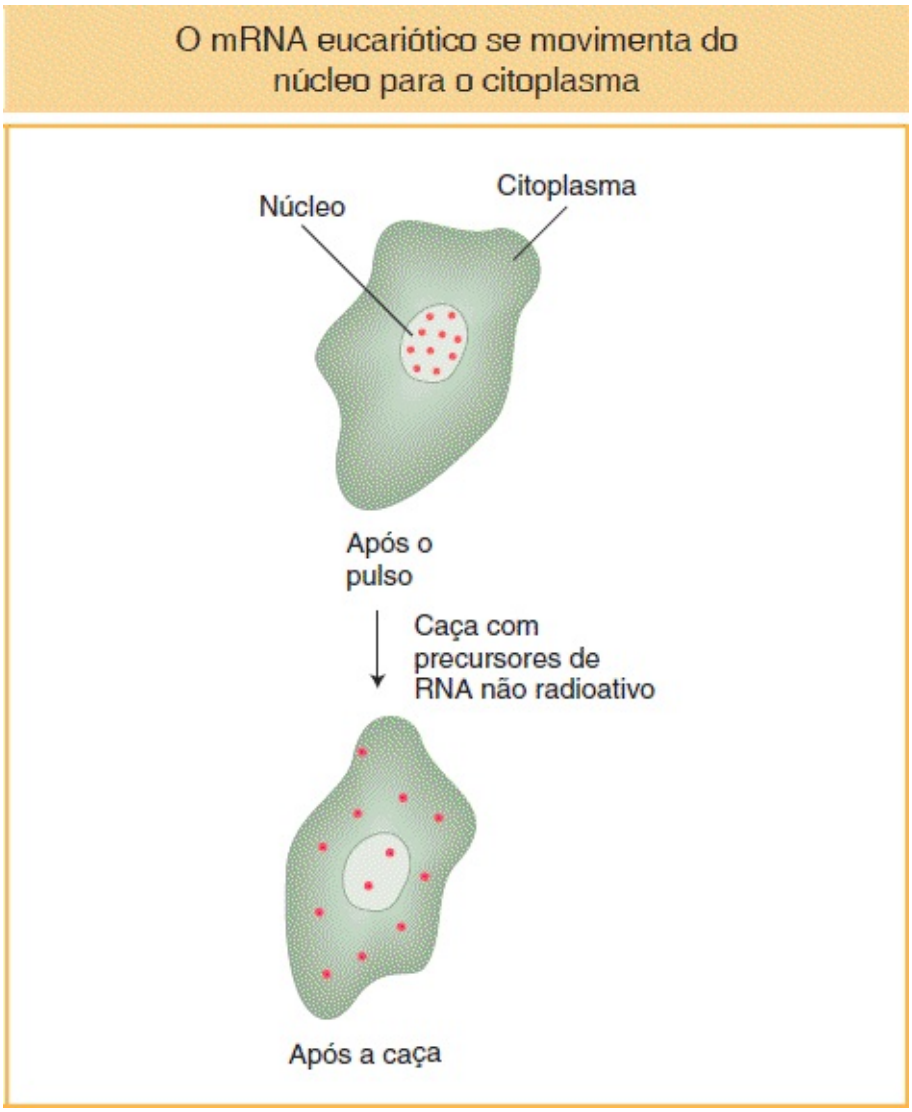
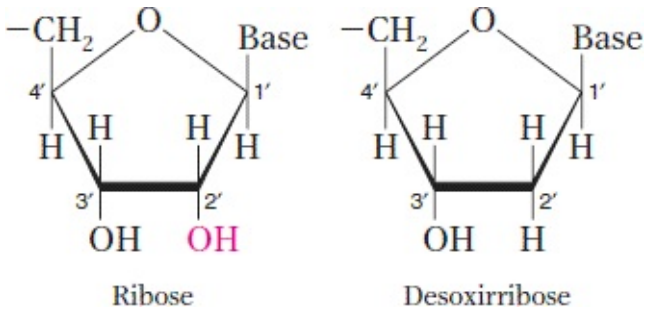


FIGURA 8.1 O experimento de pulso-caça demonstrou que o mRNA se movimenta para o citoplasma. As células são cultivadas brevemente em uracila radioativa para marcar o RNA recém-sintetizado (pulso). As células são lavadas para remover a uracila radioativa e em seguida são cultivadas em excesso de uracila não radioativa (caça). Os *pontos vermelhos* indicam a localização do RNA que contém uracila radioativa ao longo do tempo.

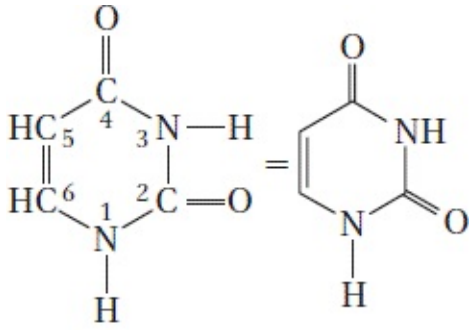


2. O RNA normalmente é uma cadeia unifilamentar de nucleotídios, não uma dupla-hélice como o DNA. Uma consequência é que o RNA é mais flexível e consegue formar uma variedade muito maior de formas moleculares tridimensionais complexas do que o DNA bifilamentar consegue. Um filamento de RNA pode dobrar-se de tal modo que algumas de suas próprias bases pareiam entre si. O referido pareamento de bases *intramolecular* é um determinante importante da forma do RNA.

Conforme você verá posteriormente neste capítulo, a presença do grupo hidroxila no átomo de carbono 2’ facilita a ação do RNA em muitos processos celulares importantes.

Assim como um filamento de DNA individual, um filamento de RNA é formado por um arcabouço de açúcar-fosfato, com uma base ligada de modo covalente à posição 1′ em cada ribose. As ligações de açúcar-fosfato são produzidas nas posições 5′ e 3′ do açúcar, justamente como no DNA; assim, uma cadeia de RNA apresentará uma extremidade 5′ e uma extremidade 3′.

3. Os nucleotídios do RNA (denominados ribonucleotídios) contêm as bases adenina, guanina e citosina, mas a base pirimidínica **uracila** (abreviada **U**) está presente em lugar da timina.



Uracila

A uracila forma duas ligações de hidrogênio com a adenina, assim como a timina. A Figura 8.2 demonstra os quatro ribonucleotídios observados no RNA.

Além disso, a uracila é capaz de parear com G. As bases U e G formam pares de bases apenas durante o dobramento do RNA, não durante a

transcrição. As duas ligações de hidrogênio que podem se formar entre U e G são mais fracas do que as duas que se formam entre U e A. A capacidade de U parear-se tanto com A quanto com G é o principal motivo pelo qual o RNA consegue formar estruturas extensas e complicadas, muitas das quais são importantes em processos biológicos.

4. O RNA — assim como as proteínas, mas contrariamente ao DNA — pode catalisar reações biológicas. A denominação **ribozima** foi cunhada para as moléculas de RNA que atuam como enzimas proteicas.

Classes de RNA

Os RNA podem ser agrupados em duas classes gerais. Uma classe de RNA codifica a informação necessária para produzir cadeias polipeptídicas (proteínas). Fazemos referência a essa classe como **RNA mensageiro (mRNA)**, tendo em vista que esses RNA, assim como um mensageiro, atuam como o intermediário que transmite a informação do DNA para a proteína. Fazemos referência à outra classe como **RNA funcional**, tendo em vista que o RNA não codifica informação para produzir proteínas. Em vez disso, o próprio RNA é o produto funcional final.

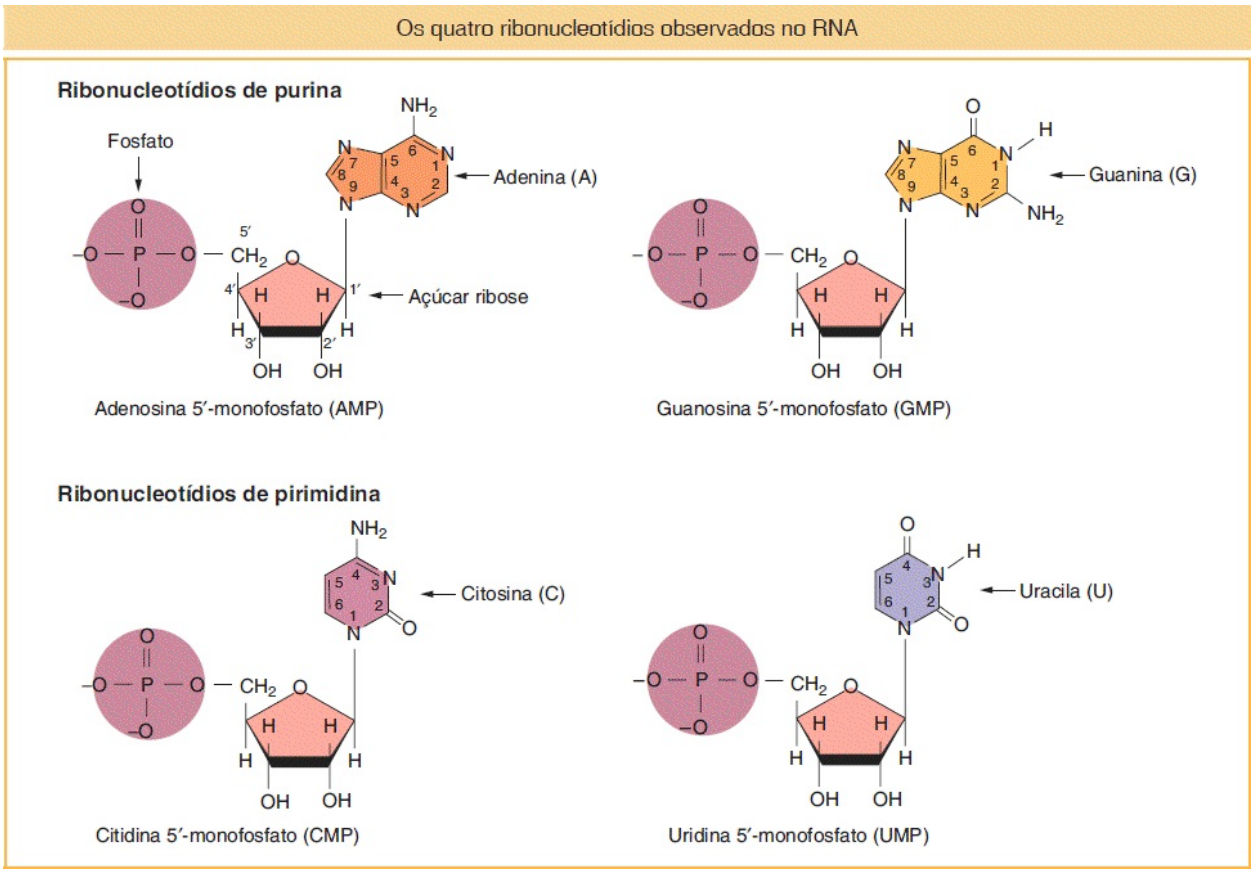


FIGURA 8.2.

RNA mensageiro. As etapas por meio das quais um gene influencia o fenótipo são denominadas **expressão gênica**. Em relação à vasta maioria dos genes, o transcrito de RNA é apenas um intermediário necessário para a síntese de uma proteína, que é o produto funcional final que influencia o fenótipo.

RNA funcional. Na medida em que mais é aprendido a respeito dos detalhes íntimos da expressão e da regulação gênica, torna-se aparente que os RNA funcionais pertencem a uma diversidade de classes e que desempenham papéis diversos. Novamente, é importante enfatizar que os RNA funcionais são ativos como RNA; eles nunca são traduzidos em polipeptídios.

As principais classes de RNA funcionais contribuem para as diversas etapas na transferência da informação do DNA para a proteína, no processamento de outros RNA e na regulação do RNA e dos níveis de proteínas na célula. Duas dessas referidas classes de RNA funcionais são observadas em procariotos e em

eucariotos: RNA transportadores e RNA ribossômicos:

- As moléculas de **RNA transportador ou de transferência (tRNA)** são responsáveis por transportar o aminoácido correto até o mRNA no processo de tradução
- As moléculas de **RNA ribossômico (rRNA)** são os principais componentes dos ribossomos, que são grandes máquinas macromoleculares que guiam a montagem da cadeia de aminoácidos pelos mRNA e tRNA.

A coleção inteira de tRNA e rRNA é codificada por um pequeno número de genes (de algumas dezenas a no máximo algumas centenas). Entretanto, embora os genes que os codifiquem sejam em pequeno número, os rRNA são responsáveis por uma porcentagem muito grande do RNA na célula, tendo em vista que são estáveis e transcritos em muitas cópias.

Outra classe de RNA funcionais participa no processamento do RNA e é específica de eucariotos:

- Os **pequenos RNA nucleares (snRNA)** são parte de um sistema que processa adicionalmente os transcritos de RNA nas células eucariotas. Alguns snRNA se unem a diversas subunidades proteicas para formar o complexo ribonucleoproteico de processamento (o *spliceossomo*) que remove os íntrons dos mRNA eucarióticos.

Finalmente, um grande grupo de RNA funcionais suprime a expressão gênica em muitos níveis e também mantém a estabilidade do genoma. Três classes desses RNA funcionais podem ser codificadas por grandes partes dos genomas eucarióticos: microRNA, pequenos RNA de interferência e RNA de interação *piwi*:

- Recentemente foi reconhecido que os **microRNA (miRNA)** apresentam um papel amplo na regulação da quantidade de proteínas produzidas por muitos genes eucarióticos
- Os **pequenos RNA de interferência (siRNA)** e os **RNA de interação *piwi* (piRNA)** auxiliam na proteção da integridade dos genomas de plantas e de

animais. Os siRNA inibem a produção de vírus, enquanto ambos, siRNA e piRNA, previnem a difusão de elementos de transposição para outros *loci* cromossômicos. Os siRNA restringem os elementos de transposição em plantas, e os piRNA realizam a mesma função em animais.

Recentemente, observou-se que os **RNA não codificadores longos** (**lncRNA**, ou por vezes abreviados apenas **ncRNA**) são transcritos a partir da maior parte das regiões dos genomas de seres humanos e de outros animais e plantas. Enquanto alguns poucos lncRNA desempenham um papel em fenômenos genéticos clássicos, tais como a compensação da dose (ver Capítulo 12), a função da maior parte dos lncRNA, se existente, atualmente é desconhecida.

Tendo em vista que a síntese proteica e o processamento de mRNA ocorrem durante todo o período de vida da maior parte das células, tRNA, rRNA e snRNA são sempre necessários. Como tal, esses RNA são sintetizados continuamente (diz-se que a sua transcrição é **constitutiva**). Contrariamente, miRNA, siRNA, piRNA e lncRNA são transcritos e/ou processados a partir de transcritos maiores de modo intermitente, apenas quando são necessários para exercer seus papéis na proteção do genoma e na regulação da expressão gênica.

> CONCEITO-CHAVE Existem duas classes gerais de RNA, aqueles que codificam proteínas (mRNA) e aqueles que são funcionais (ncRNA). Os RNA funcionais participam em uma diversidade de processos celulares, incluindo a síntese proteica (tRNA, rRNA), o processamento do RNA (snRNA), a regulação da expressão gênica (miRNA) e a defesa do genoma (siRNA, piRNA).

8.2 Transcrição

A primeira etapa na transferência da informação do gene para a proteína é produzir um filamento de RNA cuja sequência de bases seja complementar à sequência de bases de um segmento de DNA, por vezes seguida por uma

modificação daquele RNA para prepará-lo para os seus papéis celulares específicos. Portanto, o RNA é produzido por meio de um processo que copia a sequência de nucleotídios do DNA. Tendo em vista que esse processo é reminiscente da transcrição (cópia) de palavras escritas, a síntese do RNA é denominada **transcrição**. Diz-se que o DNA é transcrito em RNA, e o RNA é denominado **transcrito**.

Visão geral | O DNA como modelo da transcrição

Como a informação codificada na molécula de DNA é transferida para o transcrito de RNA? A transcrição depende do pareamento complementar de bases. Considere a transcrição de um segmento cromossômico que constitui um gene. Primeiramente, os dois filamentos da dupla-hélice de DNA são separados localmente e um dos filamentos separados atua como **molde** para a síntese de RNA. No cromossomo em geral, ambos os filamentos de DNA são utilizados como moldes, *mas, em qualquer gene, apenas um filamento é utilizado* e, naquele gene, é sempre o mesmo filamento, com início na extremidade 3′ do gene-molde (Figura 8.3). Em seguida, os ribonucleotídios que foram sintetizados quimicamente em outro local na célula formam pares estáveis com suas bases complementares no molde. O ribonucleotídio A pareia com T no DNA, G com C, C com G e U com A. Cada ribonucleotídio é posicionado oposto à sua base complementar pela enzima **RNA polimerase**. Essa enzima se liga ao DNA e se movimenta ao longo dele, ligando os ribonucleotídios alinhados para produzir uma molécula crescente de RNA, conforme demonstrado na Figura 8.4 A. Portanto, já vemos em ação os dois princípios da complementaridade das bases e da ligação ácido nucleico–proteína (nesse caso, a ligação da RNA polimerase).

Observamos que o RNA apresenta uma extremidade 5′ e uma extremidade 3′. Durante a síntese, o crescimento do RNA é sempre no sentido 5′ para 3′; em outras palavras, os nucleotídios são sempre adicionados a uma ponta em crescimento 3′, conforme demonstrado na Figura 8.4 B. Tendo em vista que os filamentos complementares de ácido nucleico estão orientados de modo oposto, o fato de o RNA ser sintetizado de 5′ para 3′ significa que o filamento-molde deve estar orientado de 3′ para 5′.

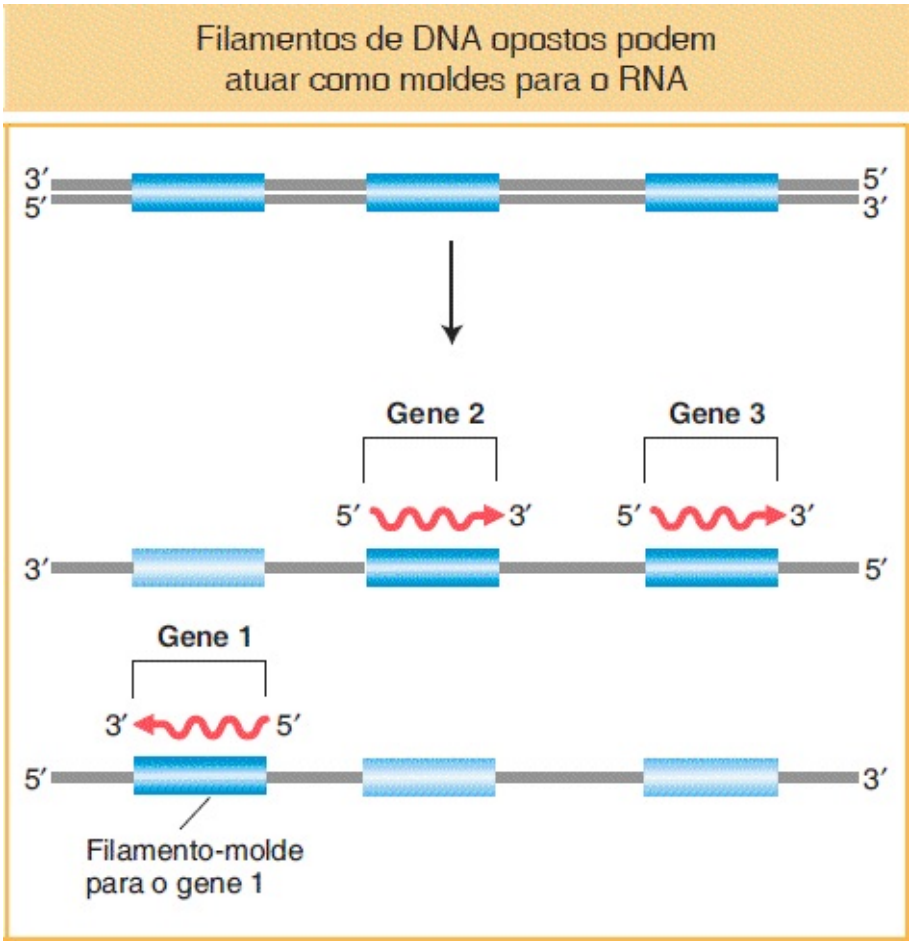


FIGURA 8.3 Apenas um filamento de DNA é o molde para a transcrição gênica, mas qual filamento varia com o gene? O sentido da transcrição é sempre o mesmo em relação a qualquer gene e tem início a partir da extremidade 3′ do molde de DNA e da extremidade 5′ do transcrito de RNA. Portanto, os genes transcritos em diferentes sentidos utilizam filamentos opostos do DNA como moldes.

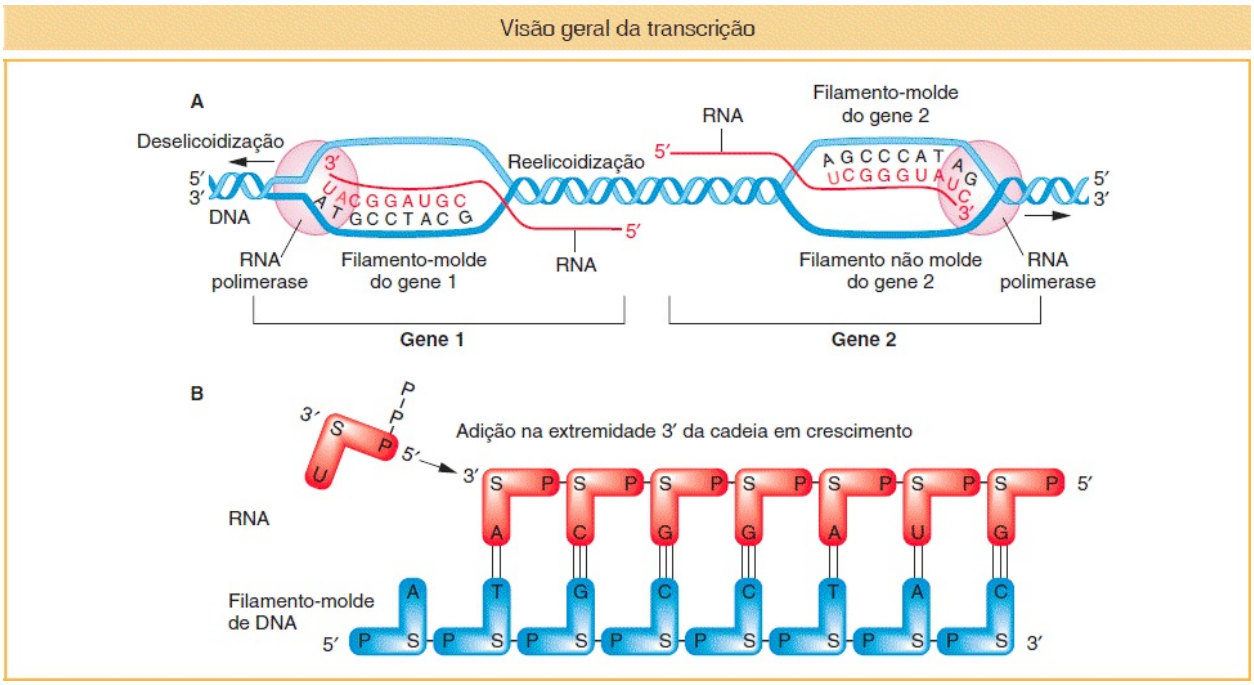


FIGURA 8.4 A. Transcrição de dois genes em sentidos opostos. Os genes 1 e 2 da Figura 8.3 estão demonstrados. O gene 1 é transcrito a partir do filamento inferior. A RNA polimerase migra para a esquerda, lendo o filamento-molde no sentido 3′ para 5′ e sintetizando RNA no sentido 5′ para 3′. O gene 2 é transcrito no sentido oposto, para a direita, tendo em vista que o filamento superior é o molde. Na medida em que a transcrição prossegue, a extremidade 5′ do RNA é deslocada do molde conforme a bolha de transcrição se fecha atrás da polimerase. **B.** Na medida em que o gene 1 é transcrito, o grupo fosfato na extremidade 5′ do ribonucleotídio (U) que entra liga-se à extremidade 3′ da cadeia de RNA em crescimento. S = Açúcar.

À medida que a molécula de RNA polimerase se movimenta ao longo do gene, ela desenrola a dupla-hélice de DNA à sua frente e enrola novamente o DNA que já foi transcrito. Conforme a molécula de RNA é progressivamente alongada, a extremidade 5′ do RNA é deslocada do molde e a bolha de transcrição se fecha atrás da polimerase. Sucessões ou moléculas sucessivas de RNA polimerases, cada uma sintetizando uma molécula de RNA, movimentam-se ao longo do gene (Figura 8.5).

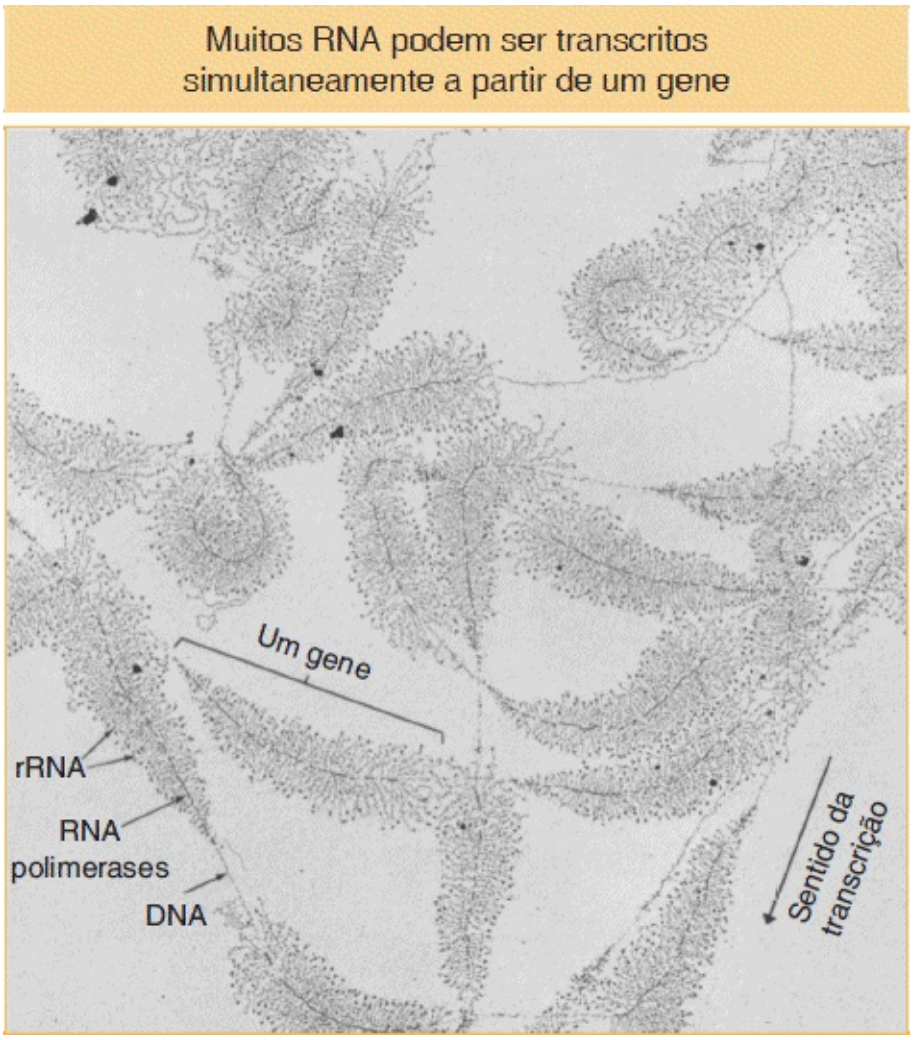


FIGURA 8.5 Esta eletromicrografia demonstra a transcrição dos genes de RNA ribossômico repetidos em *tandem* no núcleo do anfíbio *Triturus viridiscens*. Ao longo de cada gene, muitas RNA polimerases estão transcrevendo em um sentido. Os transcritos de RNA em crescimento aparecem como fios que se estendem para fora do arcabouço do DNA. Os transcritos mais curtos estão próximos do início da transcrição; os mais longos estão perto do fim do gene. O aspecto de "árvore de Natal" é o resultado. (Fotografia de O. L. Miller Jr. e Barbara A. Hamkalo.)

Também observamos que as bases no transcrito e no molde são complementares. Consequentemente, a sequência de nucleotídios no RNA deve ser a mesma daquela no filamento não molde do DNA, exceto pelo fato de que as T são substituídas por U, conforme demonstrado na Figura 8.6. Quando as sequências de bases do DNA são citadas na literatura científica, a sequência do filamento não molde é fornecida de modo convencional, tendo em vista que essa

sequência é a mesma daquela observada no RNA. Por esse motivo, o filamento não molde do DNA é denominado **filamento codificador**. É extremamente importante ter em mente essa distinção quando a transcrição é discutida.

> CONCEITO-CHAVE A transcrição é assimétrica: apenas um filamento do DNA de um gene é utilizado como um molde para a transcrição. Esse filamento está na orientação de 3′ para 5′ e o RNA é sintetizado no sentido 5′ para 3′.

Estágios da transcrição

A sequência codificadora de proteína em um gene é um segmento relativamente pequeno do DNA inserido em uma molécula de DNA muito mais longa (o cromossomo). Como o segmento apropriado é transcrito em uma molécula de RNA unifilamentar de comprimento e sequência de nucleotídios corretos? Tendo em vista que o DNA de um cromossomo é uma unidade contínua, a maquinaria de transcrição deve estar direcionada para o início de um gene para iniciar a transcrição no local certo, continuar transcrevendo o comprimento do gene, e finalmente interromper a transcrição na outra extremidade. Esses três estágios distintos da transcrição são denominados **iniciação, alongamento** e **término**. Embora o processo geral de transcrição seja notavelmente semelhante em procariotos e em eucariotos, existem diferenças importantes. Por esse motivo, acompanharemos os três estágios primeiramente em procariotos (com a utilização da bactéria intestinal *E. coli* como exemplo) e em seguida em eucariotos.

Iniciação em procariotos. Como a RNA polimerase encontra o ponto de início correto para a transcrição? Em procariotos, a RNA polimerase normalmente se liga a uma sequência de DNA específica, denominada **promotor**, localizada próximo do início da região transcrita. Um promotor é uma parte importante da região reguladora de um gene. Relembre que, tendo em vista que a síntese de um transcrito de RNA tem início na sua extremidade 5′ e continua no sentido 5′ para 3′, a convenção é desenhar e fazer referência à orientação do gene também no

sentido 5′ para 3′. Por esse motivo, o filamento de DNA não molde normalmente é demonstrado. Em geral, a extremidade 5′ é desenhada à esquerda e a 3′ à direita. Com essa visão, tendo em vista que o promotor deve estar próximo da extremidade do gene na qual a transcrição tem início, diz-se que ele está na extremidade 5′ do gene; portanto, a região promotora também é denominada região reguladora 5′ (Figura 8.7 A).

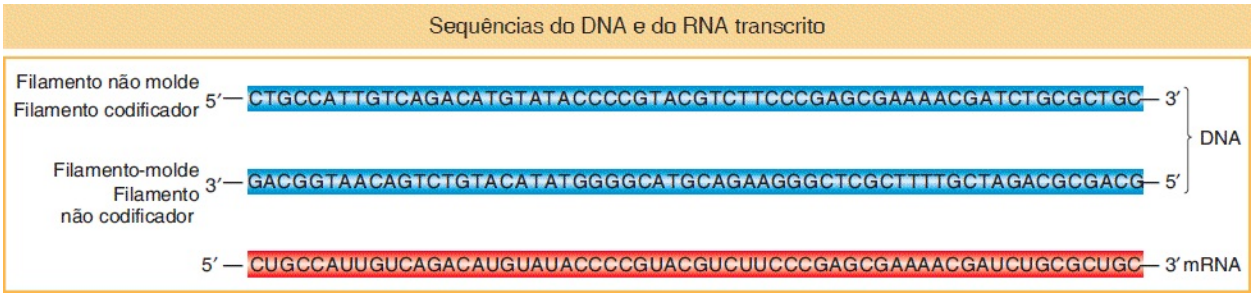


FIGURA 8.6 A sequência de mRNA é complementar ao filamento-molde de DNA a partir do qual ela é transcrita e, portanto, corresponde à sequência do filamento não molde (exceto que o RNA apresenta U onde o DNA apresenta T). Esta sequência é do gene para a enzima β-galactosidase.

A primeira base transcrita está sempre na mesma localização, designada *sítio de iniciação*. É feita referência ao promotor como ***upstream*** do sítio de iniciação, tendo em vista que está localizado à frente do sítio de iniciação (5′ do gene), no sentido oposto ao da transcrição. Um sítio ***downstream*** estaria localizado posteriormente no sentido da transcrição. Por convenção, a primeira base do DNA a ser transcrita é numerada +1. As posições dos nucleotídios *upstream* do sítio de iniciação são indicadas por um sinal negativo (—) e aquelas *dowstream*, por um sinal positivo (+).

A Figura 8.7 B demonstra as sequências promotoras de sete genes diferentes no genoma de *E. coli*. Tendo em vista que a mesma RNA polimerase se liga às sequências promotoras desses genes diferentes, as semelhanças entre os promotores não são uma surpresa. Em particular, duas regiões de grande semelhança aparecem em virtualmente todos os casos. Essas regiões foram denominadas as *regiões —35* (menos 35) e *—10*, tendo em vista que estão localizadas 35 pares de bases e 10 pares de bases, respectivamente, *upstream* da primeira base transcrita. Elas estão demonstradas em amarelo na Figura 8.7 B.

Conforme você pode observar, as regiões —35 e —10 de diferentes genes não precisam ser idênticas para realizar uma função semelhante. Não obstante, é possível chegar a uma sequência de nucleotídios, denominada **sequência consenso**, que está de acordo com a maior parte das sequências. A sequência consenso do promotor de *E. coli* está demonstrada na parte inferior da Figura 8.7 B. Uma holoenzima RNA polimerase (ver próximo parágrafo) se liga ao DNA nesse ponto, em seguida desenrola a dupla-hélice do DNA e inicia a síntese de uma molécula de RNA. Observe, na Figura 8.7 A, que a parte do gene codificadora de proteína normalmente tem início em uma sequência ATG, mas o sítio de iniciação, onde a transcrição tem início, normalmente está bem *upstream* dessa sequência. A parte intercalar é denominada **região 5′ não traduzida (5' UTR)**.

A RNA polimerase bacteriana que procura no DNA uma sequência promotora é denominada **holoenzima RNA polimerase** (Figura 8.8). Esse complexo multissubunidades é composto pelas cinco subunidades do cerne da enzima (duas subunidades α, uma β, uma β′ e uma ω), mais uma subunidade denominada **fator sigma (σ)**. As duas subunidades α auxiliam na montagem da enzima e promovem interações com proteínas reguladoras, a subunidade β é ativa na catálise, a subunidade β′ se liga ao DNA, e a subunidade ω apresenta papéis na montagem da enzima e na regulação da expressão gênica. A subunidade σ se liga às regiões —10 e —35, posicionando, assim, a holoenzima para iniciar a transcrição corretamente no sítio de início (ver Figura 8.8 A). A subunidade σ também apresenta um papel na separação (dissociação) dos filamentos de DNA ao redor da região —10, de modo que o cerne da enzima consegue se ligar fortemente ao DNA no preparo para a síntese do RNA. Após a ligação do cerne da enzima, a transcrição tem início e a subunidade σ se dissocia do restante do complexo (ver Figura 8.8 B).

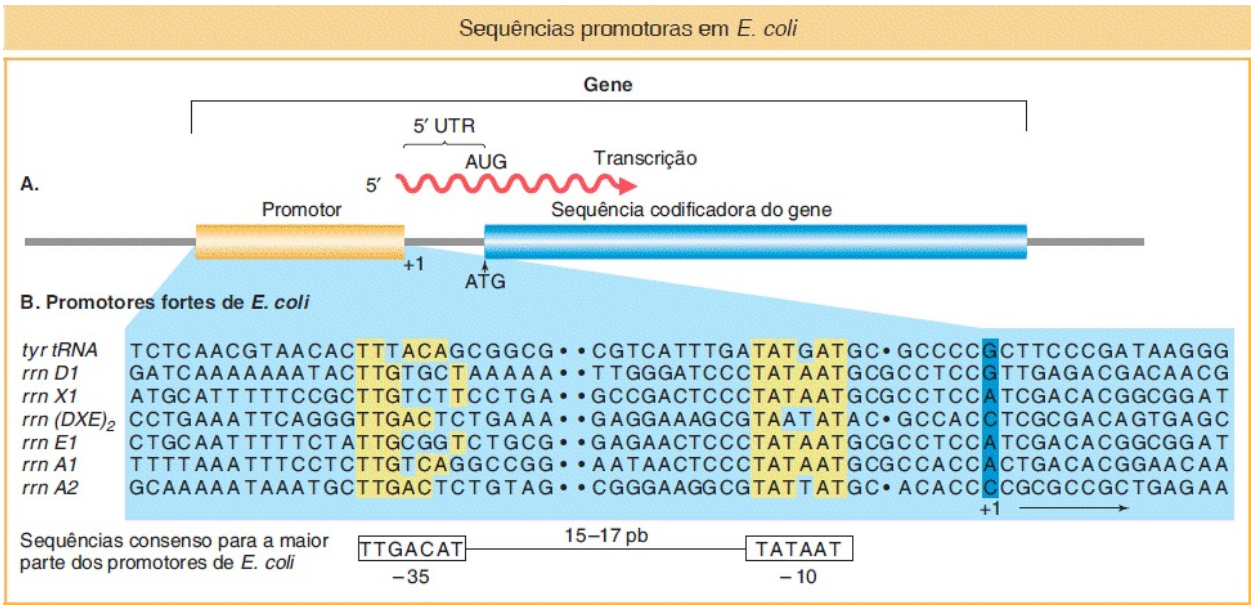


FIGURA 8.7 A. O promotor está localizado *upstream* (em direção à extremidade 5′) do ponto de início e das sequências codificadoras. **B.** Os promotores apresentam regiões de sequências semelhantes, conforme indicado pelo sombreado em *amarelo* em sete diferentes sequências promotoras em *E. coli*. Os espaços (*pontos*) estão inseridos nas sequências para otimizar o alinhamento das sequências comuns. Os números fazem referência ao número de bases antes (—) ou após (+) o ponto de iniciação da síntese de RNA. A sequência consenso em relação à maior parte dos promotores de *E. coli* está na parte inferior. (Dados de H. Lodish, D. Baltimore, A. Berk, S. L. Zipursky, P. Matsudaira e J. Darnell, *Molecular Cell Biology*, 3rd ed.)

A *E. coli*, assim como a maior parte das outras bactérias, apresenta diversos fatores σ diferentes. Um, denominado σ^{70} em virtude de sua massa em quilodáltons ser 70, é a subunidade σ primária utilizada para iniciar a transcrição da vasta maioria dos genes de *E. coli*. Outros fatores σ reconhecem diferentes sequências promotoras. Portanto, por meio da associação com diferentes fatores σ, o mesmo cerne enzimático consegue reconhecer diferentes sequências de promotor e transcrever diferentes conjuntos de genes.

Alongamento. Na medida em que a RNA polimerase se movimenta ao longo do DNA, ela desenrola o DNA à sua frente e enrola novamente o DNA que já foi transcrito. Desse modo, ela mantém uma região do DNA unifilamentar, denominada **bolha de transcrição**, dentro da qual o filamento-molde é exposto. Na bolha, a polimerase monitora a ligação de um ribonucleosídio trifosfato livre à próxima base exposta no molde de DNA e, se houver complementariedade,

adiciona-a à cadeia. A energia para a adição de um nucleotídio é derivada da divisão do trifosfato de alta energia e da liberação de difosfato inorgânico, de acordo com a fórmula geral a seguir:

$$\text{NTP} + (\text{NMP})_n \xrightarrow[\substack{\text{Mg}^{2+}\\ \text{RNA polimerase}}]{\text{DNA}} (\text{NMP})_{n+1} + PP_i$$

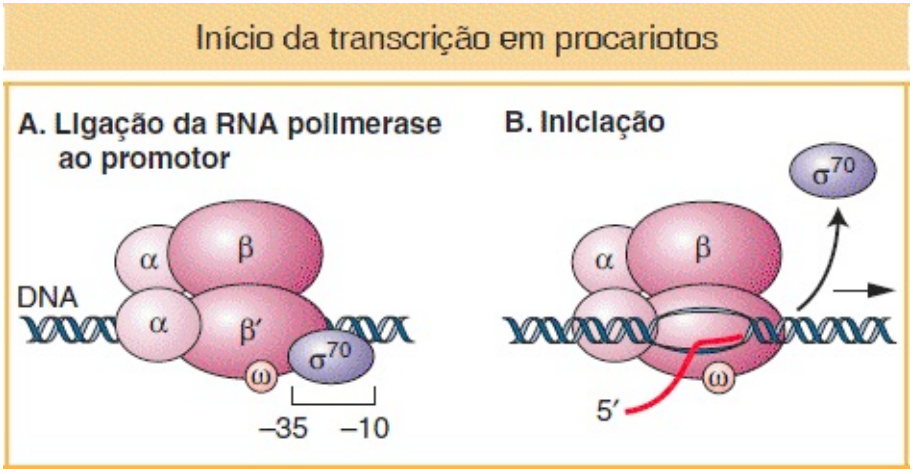


FIGURA 8.8 A subunidade σ posiciona a RNA polimerase procariótica para o início da transcrição. **A.** A ligação da subunidade σ às regiões —10 e —35 posiciona as outras subunidades para a iniciação correta. **B.** Logo após o início da síntese de RNA, a subunidade σ se dissocia das outras subunidades, que continuam a transcrição.

A Figura 8.9 A fornece uma ilustração física do alongamento. Dentro da bolha, os últimos oito ou nove nucleotídios adicionados à cadeia de RNA formam um híbrido de RNA-DNA por meio do pareamento de bases complementares com o filamento-molde. Na medida em que a cadeia de RNA é alongada em sua extremidade 3′, a extremidade 5′ é adicionalmente liberada da polimerase. Os pares de bases complementares são quebrados no ponto de saída, liberando o um filamento único.

Término. A transcrição de um gene individual continua além do segmento do gene codificador de proteína, criando uma **região 3′ não traduzida (3′ UTR)** na extremidade do transcrito. O alongamento prossegue até que a RNA polimerase reconhece sequências de nucleotídios especiais que atuam como um sinal em

relação ao término da cadeia. O encontro com os nucleotídios de sinal inicia a liberação do RNA nascente e da enzima do molde (Figura 8.9 B). Os dois mecanismos principais de término em *E. coli* (e em outras bactérias) são denominados *intrínseco* e *rho-dependente*.

No mecanismo intrínseco, o término é direto. As sequências de terminação contêm aproximadamente 40 pares de bases, terminando em um trecho rico em GC que é seguido por um trecho de seis ou mais A. Tendo em vista que G e C no molde fornecerão C e G, respectivamente, no transcrito, o RNA nessa região também é rico em GC. Essas bases C e G são capazes de formar ligações de hidrogênio complementares entre si, resultando em uma alça em grampo (Figura 8.10). Relembre que o par de bases G-C é mais estável do que o par A-T em virtude de estar ligado pelo hidrogênio em três sítos, enquanto o par A-T (ou A-U) é mantido junto por apenas duas ligações de hidrogênio. As alças em grampo do RNA, que são largamente pares G-C, são mais estáveis do que as alças que são largamente constituídas por pares A-U. A estrutura em grampo é seguida por um trecho de aproximadamente oito U que são complementares aos resíduos de A no molde de DNA.

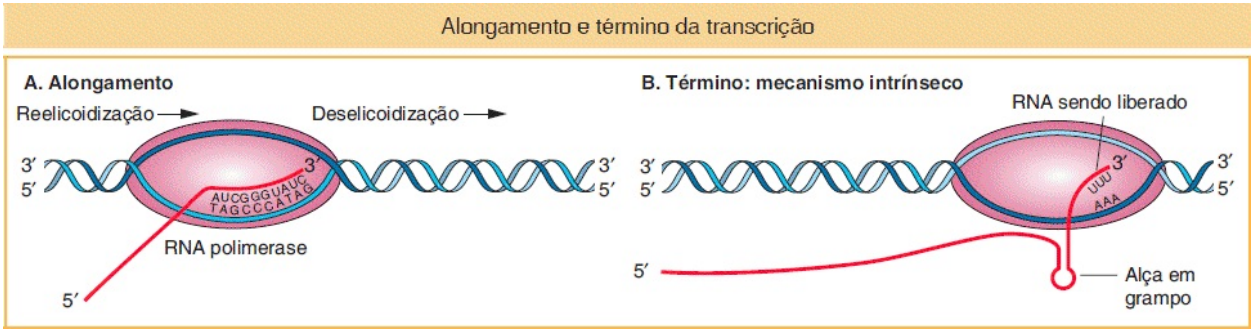


FIGURA 8.9 As cinco subunidades da RNA polimerase estão demonstradas como uma forma única semelhante a uma elipse circundando a bolha de transcrição. **A.** *Alongamento*: a síntese de um filamento de RNA complementar à região unifilamentar do filamento-molde de DNA ocorre no sentido 5′ para 3′. O DNA que é desenrolado à frente da RNA polimerase é novamente enrolado após ter sido transcrito. **B.** *Término*: o mecanismo intrínseco demonstrado aqui é um de dois modos utilizados para interromper a síntese de RNA e liberar o transcrito de RNA concluído e a RNA polimerase do DNA. Neste caso, a formação de uma alça em grampo estabelece a sua liberação. Em relação a ambos os mecanismos intrínseco e mediado por rho, o término primeiramente requer a síntese de determinadas sequências de RNA.

Normalmente, na progressão do alongamento da transcrição, a RNA polimerase irá pausar se o curto híbrido de DNA-RNA na bolha de transcrição for fraco e irá retroceder para estabilizar o híbrido. Assim como nos grampos, a força do híbrido é determinada pelo número relativo de pares de bases G-C em comparação aos pares de bases A-U (ou pares de bases A-T em híbridos de RNA-DNA). No mecanismo intrínseco, acredita-se que a polimerase pause após a síntese das U (A-U forma um híbrido de DNA-RNA fraco). Entretanto, a polimerase ao retroceder encontra a alça em grampo. Esse bloqueio no trajeto desencadeia a liberação do RNA da polimerase e a liberação da polimerase do molde de DNA.

O segundo tipo de mecanismo de término requer o auxílio de uma proteína denominada fator rho. Essa proteína reconhece as sequências de nucleotídios que atuam como sinais de término para a RNA polimerase. RNA com sinais de término rho-dependentes não apresentam o trecho de resíduos U em sua extremidade 3′ e normalmente não apresentam alças em grampo. Em vez disso, eles apresentam uma sequência de aproximadamente 40 a 60 nucleotídios que é rica em resíduos C e pobre em resíduos G e inclui um segmento *upstream* denominado sítio *rut* (*ut*ilização de *r*ho). Rho é um hexâmero composto por seis subunidades idênticas que se ligam à cadeia de RNA nascente no sítio *rut*. Esses sítios estão localizados logo *upstream* das sequências (relembre que *upstream* significa a 5′ de) nas quais a RNA polimerase tende a pausar. Após a ligação, rho facilita a liberação do RNA da RNA polimerase. Portanto o término rho-dependente envolve a ligação de rho a *rut*, a pausa da polimerase e a dissociação mediada por rho do RNA da RNA polimerase.

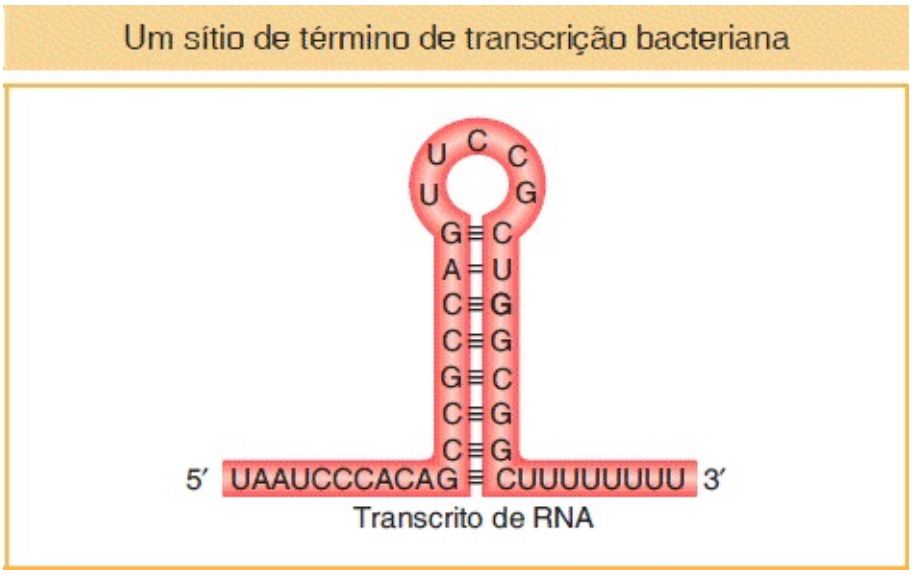


FIGURA 8.10 A estrutura de um sítio de término para a RNA polimerase em bactérias. A estrutura em grampo é formada por meio do pareamento de bases complementares em um filamento de RNA rico em GC. A maior parte dos pareamentos de bases no RNA ocorre entre G e C, mas existe um par A–U.

CONCEITO-CHAVE Os transcritos procarióticos são iniciados em 5′ da região codificadora dos genes quando a RNA polimerase se liga a uma sequência promotora consenso ou quando ela se associa a um fator σ que a guia até uma sequência promotora não consenso. O término da transcrição ocorre em sequências especiais 3′ da região codificadora que são intrínsecas ou rho-dependentes.

8.3 Transcrição em eucariotos

Conforme descrito no Capítulo 7, a replicação do DNA em eucariotos, embora mais complicada, é muito semelhante à replicação do DNA em procariotos. De algum modo, o mesmo pode ser dito em relação à transcrição, tendo em vista que eucariotos retêm muitos dos eventos associados à iniciação, ao alongamento e ao término em procariotos. A transcrição é mais complicada em eucariotos por causa de três motivos primários.

1. Os genomas eucarióticos maiores apresentam muito mais genes a ser

reconhecidos e transcritos. Enquanto as bactérias normalmente apresentam alguns milhares de genes, os eucariotos apresentam dezenas de milhares de genes. Além disso, existe muito mais DNA não codificador em eucariotos. O DNA não codificador origina-se por uma diversidade de mecanismos que serão discutidos no Capítulo 15. Assim, embora os eucariotos apresentem mais genes do que os procariotos, seus genes estão, em média, mais distantes. Por exemplo, enquanto a densidade de genes (número médio de genes por comprimento de DNA) na *E. coli* é de 1 gene por 1.400 pb, esse número cai para 1 gene por 9.000 pb na mosca-das-frutas *Drosophila*, e é de apenas 1 gene por 100.000 pb nos seres humanos. Essa baixa densidade de genes torna a etapa de iniciação da transcrição um processo muito mais complicado. Nos genomas de eucariotos multicelulares, encontrar o início de um gene pode ser como encontrar uma agulha em um palheiro.

Como você verá, os eucariotos lidam com essa situação de diversos modos. Primeiramente, eles dividiram o trabalho da transcrição entre três polimerases diferentes:

a. A RNA polimerase I transcreve os genes de rRNA (excluindo o rRNA 5S)

b. A RNA polimerase II transcreve todos os genes codificadores de proteínas, em relação aos quais o transcrito final é o mRNA, e transcreve alguns snRNA

c. A RNA polimerase III transcreve os pequenos genes de RNA funcional (tais como os genes para tRNA, alguns snRNA e o rRNA 5S).

Nesta seção, focaremos nossa atenção na RNA polimerase II.

Em segundo lugar, os eucariotos necessitam da montagem de muitas proteínas em um promotor antes que a RNA polimerase II possa começar a sintetizar o RNA. Algumas dessas proteínas, denominadas **fatores gerais de transcrição** (**GTF**; do inglês, *general transcription factors*), ligam-se antes da ligação da RNA polimerase II, enquanto outras se ligam posteriormente. O papel dos GTF e sua interação com a RNA polimerase II serão descritos na próxima seção, sobre o início da transcrição em eucariotos.

2. Uma diferença significativa entre eucariotos e procariotos é a presença de

um núcleo em eucariotos. Em procariotos, que apresentam ausência de uma membrana nuclear, a informação no RNA é quase imediatamente traduzida em uma cadeia de aminoácidos (polipeptídios), conforme veremos no Capítulo 9. Em eucariotos, a transcrição e a tradução são espacialmente separadas — a transcrição ocorre no núcleo e a tradução no citoplasma (Figura 8.11). Em eucariotos, o RNA é sintetizado no núcleo onde o DNA está localizado e é exportado do núcleo para o citoplasma para tradução.

Antes que o RNA deixe o núcleo, ele deve ser modificado de diversos modos. Essas modificações são coletivamente denominadas **processamento do RNA**. Para distinguir o RNA antes e após o processamento, o RNA recém-sintetizado é denominado **transcrito primário** ou **pré-mRNA**, e o termo *mRNA* é reservado para o transcrito totalmente processado que pode ser exportado para fora do núcleo. Conforme veremos, a extremidade 5′ do RNA é submetida ao processamento enquanto a extremidade 3′ ainda está sendo sintetizada. Portanto, a RNA polimerase II deve sintetizar o RNA enquanto coordena simultaneamente um arranjo diverso de eventos de processamento. Por esse motivo, entre outros, a RNA polimerase II é uma enzima formada por multissubunidades, mais complexa que a RNA polimerase procariótica. Na verdade, ela é considerada outra máquina molecular. A coordenação do processamento e da síntese do RNA pela RNA polimerase II será discutida na seção sobre o alongamento da transcrição em eucariotos.

3. Finalmente, o molde para a transcrição, o DNA genômico, está organizado em cromatina nos eucariotos (ver Capítulo 1), enquanto está virtualmente "despido" em procariotos. Como você aprenderá no Capítulo 12, determinadas estruturas da cromatina podem bloquear o acesso da RNA polimerase ao molde de DNA. Essa característica da cromatina se desenvolveu em um mecanismo muito sofisticado para a regulação da expressão gênica em eucariotos. Entretanto, uma discussão sobre a influência da cromatina na capacidade da RNA polimerase II de iniciar a transcrição será colocada à margem até o Capítulo 12, tendo em vista que focamos nos eventos que ocorrem *após* a RNA polimerase II ter o acesso ao molde de DNA.

Iniciação da transcrição em eucariotos

Como declarado anteriormente, a transcrição tem início em procariotos quando a subunidade σ da holoenzima RNA polimerase reconhece as regiões —10 e —35 no promotor de um gene. Após o início da transcrição, a subunidade σ se dissocia e o cerne da polimerase continua a sintetizar o RNA dentro de uma bolha de transcrição que se movimenta ao longo do DNA. De modo semelhante, em eucariotos, o cerne da RNA polimerase II também não consegue reconhecer sequências promotoras por si próprio. Entretanto, contrariamente às bactérias, nas quais o fator σ é uma parte integrante da holoenzima polimerase, os eucariotos necessitam de GTF para se ligar às regiões no promotor *antes* da ligação do cerne da enzima.

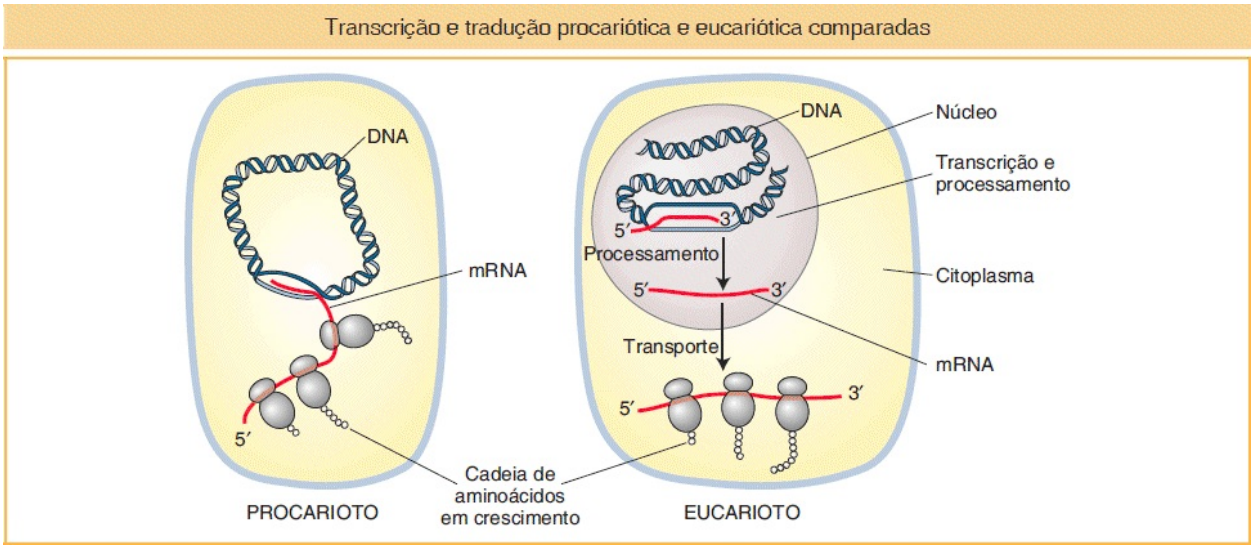


FIGURA 8.11 A transcrição e a tradução ocorrem no mesmo compartimento celular em procariotos, mas em compartimentos diferentes nos eucariotos. Além disso, contrariamente aos transcritos de RNA procarióticos, os transcritos eucarióticos são submetidos a um extensivo processamento antes que possam ser traduzidos em proteínas.

O início da transcrição em eucariotos apresenta algumas características que são reminiscentes do início da replicação do DNA nas origens de replicação. Relembre do Capítulo 7 que proteínas que não são parte do replissomo iniciam a montagem da máquina de replicação. A proteína DnaA em *E. coli* e o complexo de reconhecimento da origem (ORC) em leveduras, por exemplo, primeiramente

reconhecem e se ligam à origem nas sequências de DNA. Essas proteínas atuam para atrair proteínas de replicação, incluindo a DNA polimerase III, por meio de interações proteína–proteína. De modo semelhante, GTF, que não participam da síntese de RNA, reconhecem e se ligam às sequências no promotor ou a outros GTF e atuam para atrair o cerne da RNA polimerase II e posicioná-lo no sítio correto a fim de iniciar a transcrição. Os GTF são designados TFIIA, TFIIB e assim por diante (em referência a *t*ranscription *factor* of RNA polymerase *II*, fator de transcrição da RNA polimerase II, em inglês).

Os GTF e o cerne da RNA polimerase II constituem o **complexo de pré-iniciação** (**PIC**; do inglês, *preinitiation complex*). Esse complexo é consideravelmente grande: ele contém seis GTF, *cada um* dos quais é um complexo multiproteico, mais o cerne da RNA polimerase II, que é composto por uma dúzia ou mais de subunidades proteicas. A sequência de aminoácidos de algumas das subunidades do cerne da RNA polimerase II é conservada desde leveduras até seres humanos. Essa conservação pode ser dramaticamente demonstrada por meio da substituição de algumas subunidades de RNA polimerase II de levedura por suas correspondentes humanas para formar um complexo de RNA polimerase II *quimérico* (denominado em referência a uma criatura da mitologia grega que lançava fogo e que apresentava uma cabeça de leão, um corpo de cabra e uma cauda de serpente). Esse complexo de RNA polimerase II quimérico é totalmente funcional em leveduras.

Assim como os promotores procarióticos, os promotores eucarióticos estão localizados no lado 5′ (*upstream*) do sítio de início da transcrição. Quando regiões promotoras eucarióticas de diferentes espécies são alinhadas, com frequência pode-se observar que a sequência TATA está localizada a aproximadamente 30 pares de bases (—30 pb) do local de início da transcrição (Figura 8.12). Essa sequência, denominada **TATA boxe**, é o sítio do primeiro evento na transcrição: a ligação da **proteína de ligação a TATA** (**TBP**; do inglês, *TATA-binding protein*). A TBP é parte do complexo TFIID, um dos seis GTF. Quando ligada ao TATA boxe, a TBP atrai outros GRF e o cerne da RNA polimerase II para o promotor, formando, assim, o PIC. Após o início da transcrição, a RNA polimerase II se dissocia da maior parte dos GTF para

alongar o transcrito primário de RNA. Alguns dos GTF permanecem no promotor para atrair o próximo cerne de RNA polimerase. Assim, diversas enzimas RNA polimerase II podem sintetizar simultaneamente transcritos de um único gene.

Como o cerne da RNA polimerase II é capaz de se separar dos GTF e iniciar a transcrição? Embora os detalhes desse processo ainda estejam sendo analisados, sabe-se que a subunidade β da RNA polimerase II contém uma cauda proteica, denominada **domínio carboxi-terminal** (**CTD**; do inglês, *carboxy terminal domain*), que desempenha um papel-chave. O CTD está estrategicamente localizado próximo ao sítio no qual o RNA nascente surgirá a partir da polimerase. A fase de iniciação termina e a fase de alongamento tem início após o CTD ter sido fosforilado por um dos GTF. Acredita-se que essa fosforilação de algum modo enfraqueça a ligação da RNA polimerase II com outras proteínas do PIC e possibilite o alongamento. O CTD também participa em diversas outras fases críticas da síntese e do processamento do RNA.

Iniciação da transcrição em eucariotos

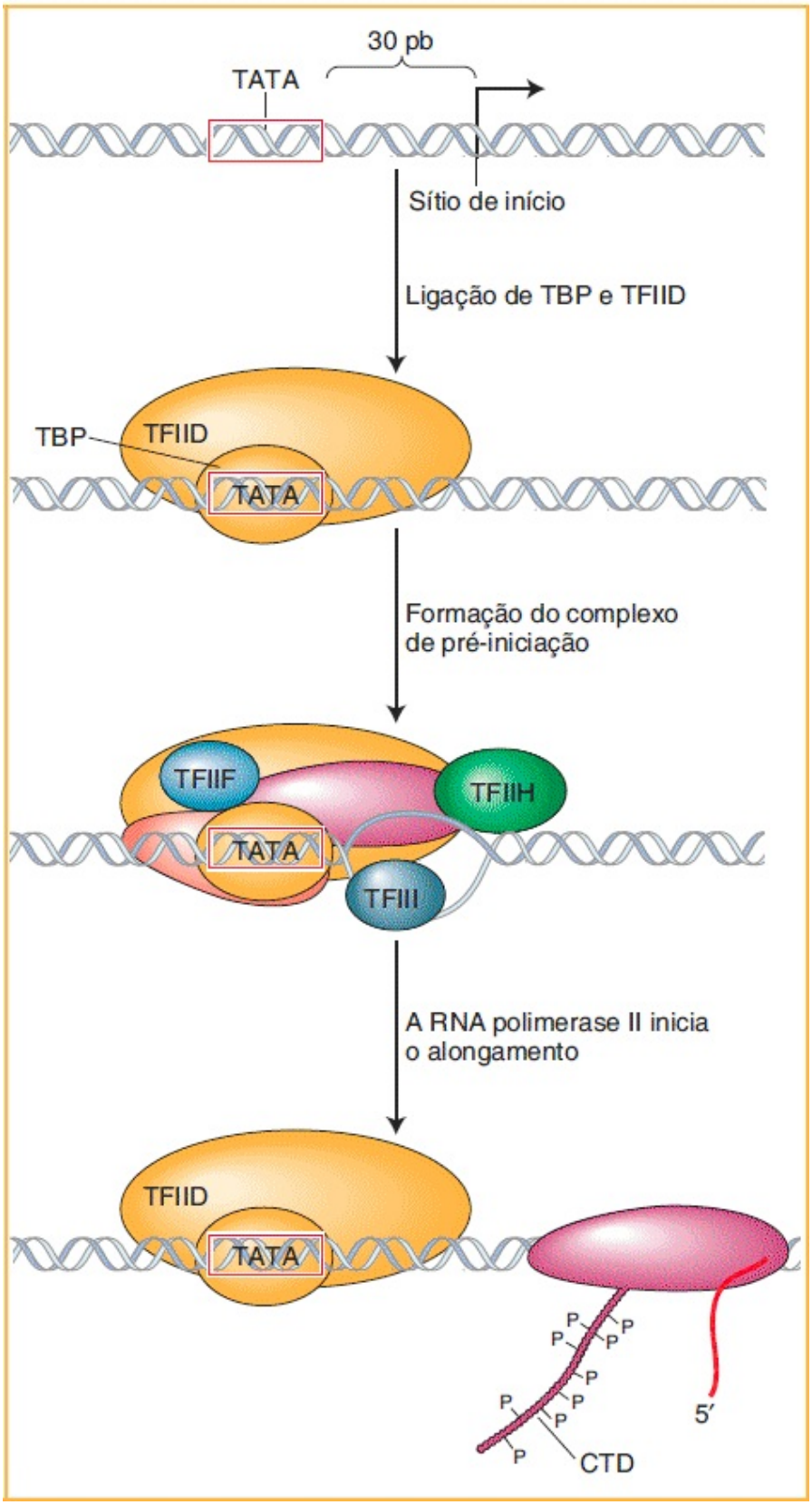


FIGURA 8.12 A formação do complexo de pré-iniciação normalmente tem início com a ligação da proteína de ligação a TATA (TBP), que em seguida recruta os outros fatores gerais de transcrição (GTF) e a RNA polimerase II para o sítio de início da transcrição. A transcrição tem início após a fosforilação do domínio carboxi-terminal (CTD) da RNA polimerase II.

> CONCEITO-CHAVE Os promotores eucarióticos são reconhecidos primeiramente pelos fatores gerais de transcrição (GTF). A função dos GTF é atrair o cerne da RNA polimerase II, de modo que ela seja posicionada para iniciar a síntese de RNA no sítio de início da transcrição.

Alongamento, término e processamento de pré-mRNA em eucariotos

O alongamento ocorre dentro da bolha de transcrição essencialmente conforme descrito em relação à síntese do RNA procariótico. Entretanto, o RNA nascente apresenta destinos muito diferentes em procariotos e em eucariotos. Em procariotos, a tradução tem início na extremidade 5′ do RNA nascente, enquanto a metade 3′ ainda está sendo sintetizada. Contrariamente, o RNA dos eucariotos deve ser submetido a um processamento adicional antes que ele possa ser traduzido. Esse processamento inclui (1) a adição de um revestimento (*cap*) na extremidade 5′, (2) a recomposição (*splicing*) para eliminar os íntrons e (3) a adição de uma cauda de nucleotídios adenina (poliadenilação) em 3′.

Assim como a replicação do DNA, a síntese e o processamento do pré-mRNA em mRNA requerem que muitas etapas sejam realizadas rapidamente e com precisão. Primeiramente, acreditava-se que a maior parte do processamento do pré-mRNA eucariótico ocorresse após a conclusão da síntese do RNA. O processamento após a conclusão da síntese do RNA é denominado **pós-transcricional**. Entretanto, atualmente evidências experimentais indicam que o processamento realmente ocorre durante a síntese de RNA; ele é **cotranscricional**. Portanto, o RNA parcialmente sintetizado (nascente) está sendo submetido a reações de processamento na medida em que emerge do complexo da RNA polimerase II.

O CTD da RNA polimerase II eucariótica desempenha um papel central na

coordenação de todos os eventos de processamento. O CTD é composto por muitas repetições de uma sequência de sete aminoácidos. Essas repetições atuam como sítios de ligação para algumas das enzimas e outras proteínas que são necessárias para o revestimento do RNA, a recomposição e a clivagem seguida pela poliadenilação. O CTD está localizado próximo do sítio onde o RNA nascente emerge da polimerase e, assim, está em um local ideal para orquestrar a ligação e a liberação das proteínas necessárias para processar o transcrito de RNA nascente enquanto a síntese do RNA continua. Nas diversas fases do processamento, os aminoácidos do CTD são modificados de modo reversível — normalmente por meio da adição e da remoção de grupos fosfato (denominadas fosforilação e desfosforilação, respectivamente). O estado de fosforilação do CTD determina que proteínas de processamento podem se ligar. Desse modo, o CTD determina a tarefa a ser realizada no RNA na medida em que ele emerge da polimerase. Os eventos de processamento e o papel do CTD na sua execução estão demonstrados na Figura 8.13 e são considerados em seguida.

Processamento cotranscricional do RNA

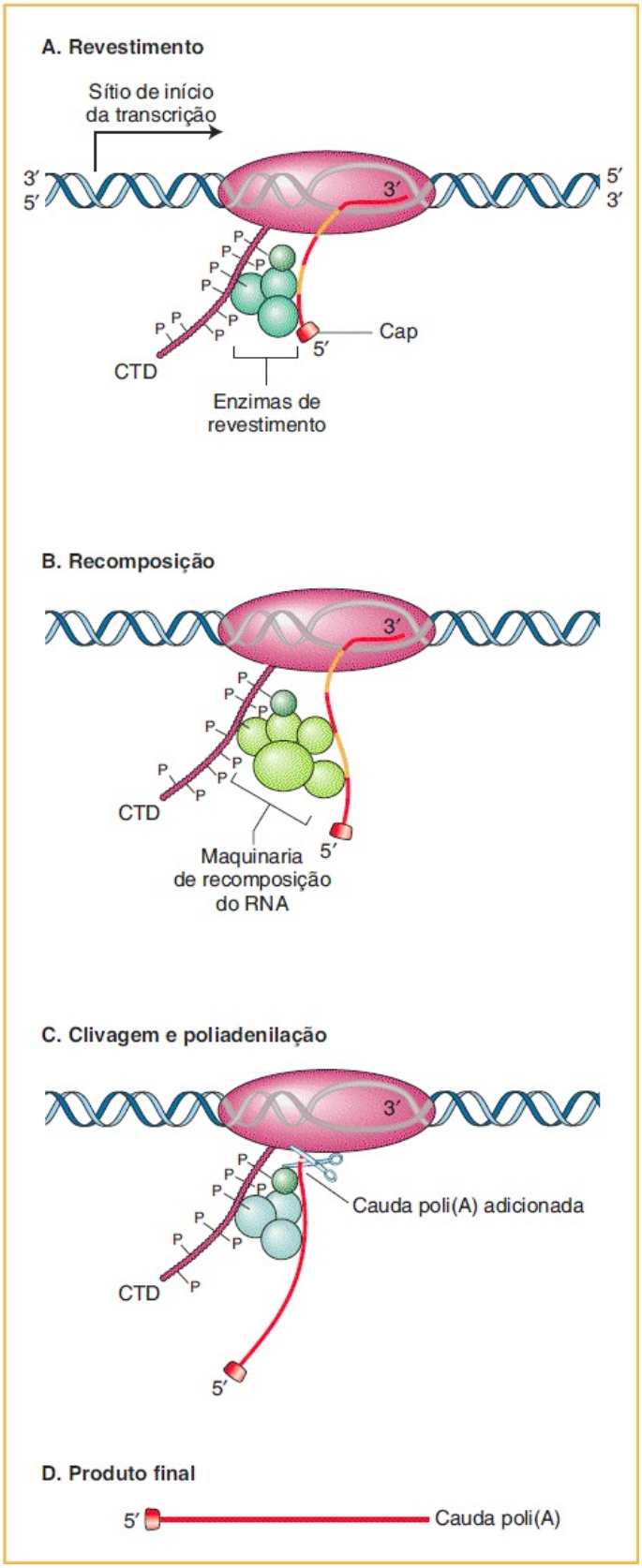


FIGURA 8.13 O processamento cotranscricional do RNA é coordenado pelo domínio carboxi-terminal (CTD) da subunidade β da RNA polimerase II. A fosforilação reversível dos aminoácidos de CTD (indicada por P) cria sítios de ligação para as diferentes enzimas de processamento e os fatores necessários para **(A)** revestimento, **(B)** recomposição, **(C)** clivagem e poliadenilação.

Processamento das extremidades 5′ e 3′. A Figura 8.13 A ilustra o processamento da extremidade 5′ do transcrito de um gene codificador de proteína. Quando o RNA nascente emerge pela primeira vez a partir da RNA polimerase II, uma estrutura especial, denominada ***cap* (revestimento)**, é adicionada à extremidade 5′ por diversas proteínas que interagem com o CTD. O *cap* consiste em um resíduo 7-metilguanosina ligado ao transcrito por três grupos fosfato. O *cap* apresenta duas funções. Primeiramente, ele protege o RNA da degradação em sua longa jornada até o local da tradução. Em segundo lugar, conforme você verá no Capítulo 9, o *cap* é necessário para a tradução do mRNA.

O alongamento do RNA continua até que a sequência conservada AAUAAA ou AUUAAA seja alcançada, marcando a extremidade 3′ do transcrito. Uma enzima reconhece aquela sequência e corta a extremidade do RNA a aproximadamente 20 bases adiante. A essa extremidade cortada, é adicionado um trecho de 150 a 200 nucleotídios adenina, denominado **cauda poli(A)** (ver Figura 8.13 C). Portanto, a sequência AAUAAA do mRNA de genes codificadores de proteínas é denominada *sinal de poliadenilação*.

Recomposição do RNA, a remoção de íntrons. Em 1977, surgiu um estudo científico intitulado “An amazing sequence arrangement at the 5′ end of adenovirus 2 messenger RNA” [Um arranjo incrível de sequência na extremidade 5’ do RNA mensageiro do adenovírus 2].[1] Os cientistas normalmente suavizam suas palavras, pelo menos em suas publicações, e a utilização da palavra *amazing* [incrível] indicou que algo verdadeiramente inesperado havia sido observado. Os laboratórios de Richard Roberts e Phillip Sharp haviam descoberto, de modo independente, que a informação codificada pelos genes eucarióticos (no caso, o gene de um vírus que infecta células eucarióticas) pode ser fragmentada em segmentos de dois tipos: éxons e íntrons. Conforme declarado

anteriormente, os segmentos que codificam partes das proteínas são os éxons e os segmentos que separam os éxons são os íntrons. Os íntrons estão presentes não apenas em genes codificadores de proteínas, mas também em alguns rRNA e até mesmo em genes de tRNA.

Os íntrons são removidos do transcrito primário enquanto o RNA ainda está sendo transcrito e após o *cap* ter sido adicionado, mas antes de o transcrito ser transportado para o citoplasma. A remoção dos íntrons e a união dos éxons é denominada **recomposição (*splicing*)**, tendo em vista que é reminiscente do modo pelo qual uma fita de vídeo ou um filme podem ser cortados e reunidos para deletar um segmento específico. A recomposição une as regiões codificadoras, ou éxons, de modo que o mRNA agora contém uma sequência codificadora que é completamente colinear à proteína que ela codifica.

O número e o tamanho dos íntrons varia de gene para gene e de espécie para espécie. Por exemplo, apenas 200 dos 6.300 genes em leveduras apresentam íntrons, enquanto genes típicos em mamíferos, incluindo seres humanos, apresentam diversos. O tamanho médio de um íntron de mamífero é de aproximadamente 2.000 nucleotídios e um éxon médio apresenta aproximadamente 200 nucleotídios; portanto, uma porcentagem maior de DNA em mamíferos codifica íntrons em vez de éxons. Um exemplo extremo é o gene humano da distrofia muscular de Duchenne. Esse gene apresenta 79 éxons e 78 íntrons em seus 2,5 milhões de pares de bases. Quando submetidos à recomposição, seus 79 éxons produzem um mRNA de 14.000 nucleotídios, o que significa que os íntrons são responsáveis pela vasta maioria dos 2,5 milhões de pares de bases.

Recomposição alternativa. Nesse ponto, você pode estar ponderando a respeito da utilidade de ter genes organizados em éxons e íntrons. Relembre que este capítulo teve início com uma discussão sobre o número de genes no genoma humano. Esse número (agora estimado em aproximadamente 21.000 genes) é inferior ao dobro da quantidade de genes em nematódeos, ainda que o espectro de proteínas humanas (denominado **proteoma**; ver Capítulo 9) seja superior a 70.000. O fato de as proteínas superarem tanto o número de genes indica que um

gene pode codificar informação para mais de uma proteína. Um modo como um gene pode codificar múltiplas proteínas é por meio de um processo denominado **recomposição alternativa** (ou *splicing* alternativo). Nesse processo, diferentes mRNAs e, subsequentemente, diferentes proteínas são produzidos a partir do mesmo transcrito primário por meio da recomposição de diferentes combinações de éxons. Por motivos que atualmente são desconhecidos, a proporção de genes submetidos à recomposição alternativa varia de espécie para espécie. Embora a recomposição alternativa seja rara em plantas, mais de 70% dos genes humanos são recompostos de maneira alternativa. Muitas mutações com consequências sérias para o organismo ocorrem em virtude de defeitos na recomposição.

As consequências da recomposição alternativa sobre a estrutura e a função proteica serão apresentadas posteriormente no livro. Por enquanto, é suficiente dizer que as proteínas produzidas por meio da recomposição alternativa normalmente estão relacionadas (tendo em vista que elas normalmente contêm subconjuntos dos mesmos éxons do transcrito primário) e que com frequência elas são utilizadas em diferentes tipos celulares ou em diferentes estágios do desenvolvimento. A Figura 8.14 demonstra a miríade de combinações produzidas por meio da recomposição alternativa do transcrito primário de RNA do gene da α-tropomiosina. O mecanismo de recomposição é considerado na próxima seção.

> CONCEITO-CHAVE O pré-mRNA eucariótico é extensivamente processado antes de ser transportado como mRNA até o citoplasma para a tradução em proteína. São adicionados um *cap* em 5′ e uma cauda poli(A) em 3′; os íntrons são removidos e os éxons são unidos. Um gene pode codificar mais de um polipeptídio quando o seu pré-mRA é submetido à recomposição alternativa.

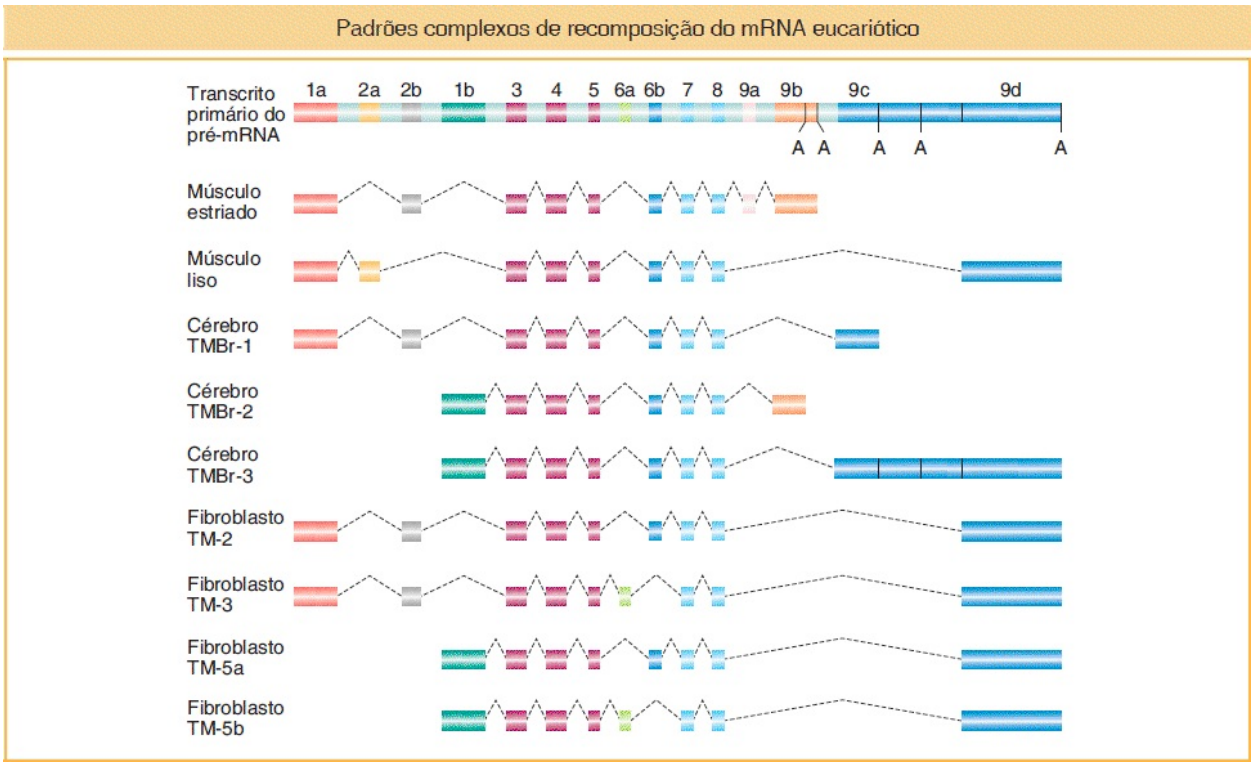


FIGURA 8.14 O pré-mRNA transcrito do gene da α-tropomiosina de rato é submetido à recomposição alternativa em diferentes tipos celulares. Os campos *verde-claros* representam íntrons; as outras cores representam éxons. Os sinais de poliadenilação estão indicados por um A. As *linhas tracejadas* nos mRNA maduros indicam regiões que foram removidas pela recomposição. TM = Tropomiosina. (Dados de J. P. Lees *et al.*, *Mol. Cell. Biol.* 10, 1990, 1729-1742.)

8.4 Remoção de íntrons e recomposição de éxons

Em virtude de o RNA ser uma molécula tão versátil, ele participa em uma diversidade de processos celulares. No Capítulo 9, você aprenderá mais a respeito do papel dos RNA funcionais como importantes componentes do ribossomo, a máquina biológica que é o local da síntese proteica. Nesta seção e na próxima, você verá que os RNA funcionais também apresentam papéis proeminentes no processamento do mRNA e na regulação de seu nível na célula.

Pequenos RNA nucleares (snRNA) | O mecanismo de recomposição de éxons

Após a descoberta dos éxons e dos íntrons, cientistas voltaram a sua atenção para

o mecanismo de recomposição do RNA. Tendo em vista que os íntrons precisam ser removidos com precisão e que os éxons precisam ser unidos com precisão, a primeira abordagem foi comparar as sequências de pré-mRNA para indicações sobre como os íntrons e os éxons são reconhecidos. A Figura 8.15 demonstra as junções éxon–íntron dos pré-mRNA. Essas junções são os sítios nos quais ocorrem as reações de recomposição. Nessas junções, observou-se que determinados nucleotídios específicos são quase idênticos entre os genes e entre as espécies; eles foram altamente conservados em virtude de participarem nas reações de recomposição. Cada íntron é cortado em cada extremidade, e essas extremidades do íntron quase sempre apresentam GU na extremidade 5′ e AG na extremidade 3′ (a **regra GU–AG**). Outro sítio invariante é um resíduo A (o ponto de ramificação A) entre 15 e 45 nucleotídios *upstream* do sítio de corte 3′. Os nucleotídios que flanqueiam os nucleotídios altamente conservados também são conservados, mas em menor grau. A existência de sequências de nucleotídios conservadas nas junções de corte sugeriu que deve haver uma maquinaria celular que reconhece essas sequências e realiza a recomposição. Assim como em geral ocorre na pesquisa científica, a maquinaria de recomposição foi encontrada acidentalmente e o mecanismo da recomposição era totalmente inesperado.

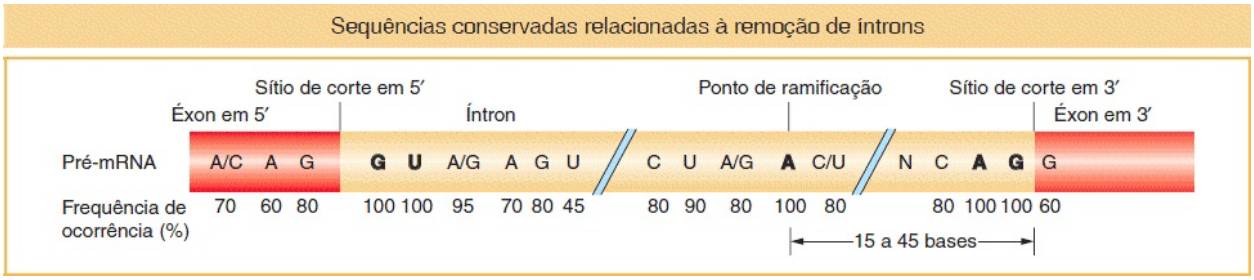


FIGURA 8.15 Sequências de nucleotídios conservados estão presentes nas junções de íntrons e éxons. Os números abaixo dos nucleotídios indicam a porcentagem de semelhança entre os organismos. De particular importância são os resíduos G e U na extremidade 5′, e os resíduos A e G na extremidade 3′, e o resíduo A rotulado “ponto de ramificação” (ver Figura 8.17 para uma visão da estrutura da ramificação). N representa qualquer base.

Um achado afortunado no laboratório de Joan Steitz levou à descoberta dos componentes da maquinaria de recomposição. Pacientes com uma diversidade de

doenças autoimunes, incluindo lúpus eritematoso sistêmico, produzem anticorpos contra suas próprias proteínas. Durante a análise de amostras sanguíneas de pacientes com lúpus, Steitz e colegas identificaram anticorpos que conseguiam se ligar a um grande complexo molecular de pequenos RNA e proteínas. Em virtude de esse complexo riboproteico estar localizado no núcleo, os componentes do RNA foram denominados *pequenos RNA nucleares*. Observou-se que os snRNA são complementares às sequências consenso nas junções de recomposição, levando os cientistas a formularem a hipótese de um papel para os snRNA no processo de recomposição. Os nucleotídios conservados no transcrito sabidamente são reconhecidos por cinco pequenas ribonucleoproteínas nucleares (snRNP), que são complexos de proteínas e um de cinco snRNA (U1, U2, U4, U5 e U6). Essas snRNP e mais de 100 proteínas adicionais são parte do spliceossomo, a grande máquina biológica que remove os íntrons e une os éxons. Os componentes do spliceossomo interagem com o CTD, conforme sugerido na Figura 8.13 B.

Esses componentes do spliceossomo se unem às sequências de íntrons e éxons, conforme demonstrado na Figura 8.16. As snRNP U1 e U2 ajudam a alinhar os sítios de corte em qualquer extremidade de um íntron, por meio da formação de ligações de hidrogênio com o íntron conservado, e as sequências de éxons. Em seguida os snRNP recrutam U4, U5 e U6 e formam o spliceossomo, que catalisa a remoção do íntron por meio de duas etapas consecutivas de recomposição (ver Figura 8.16). A primeira etapa une uma extremidade do íntron à adenina interna conservada, formando uma estrutura em alça, no formato de um laço de caubói. A segunda etapa libera o laço e une os dois éxons adjacentes. A Figura 8.17 retrata a química por trás da remoção do íntron e da recomposição dos éxons. Quimicamente, as duas etapas são reações de transesterificação entre os nucleotídios conservados. Os grupos hidroxila nas posições 2′ e 3′ dos ribonucleotídios são participantes-chave da reação.

Montagem e função do spliceossomo

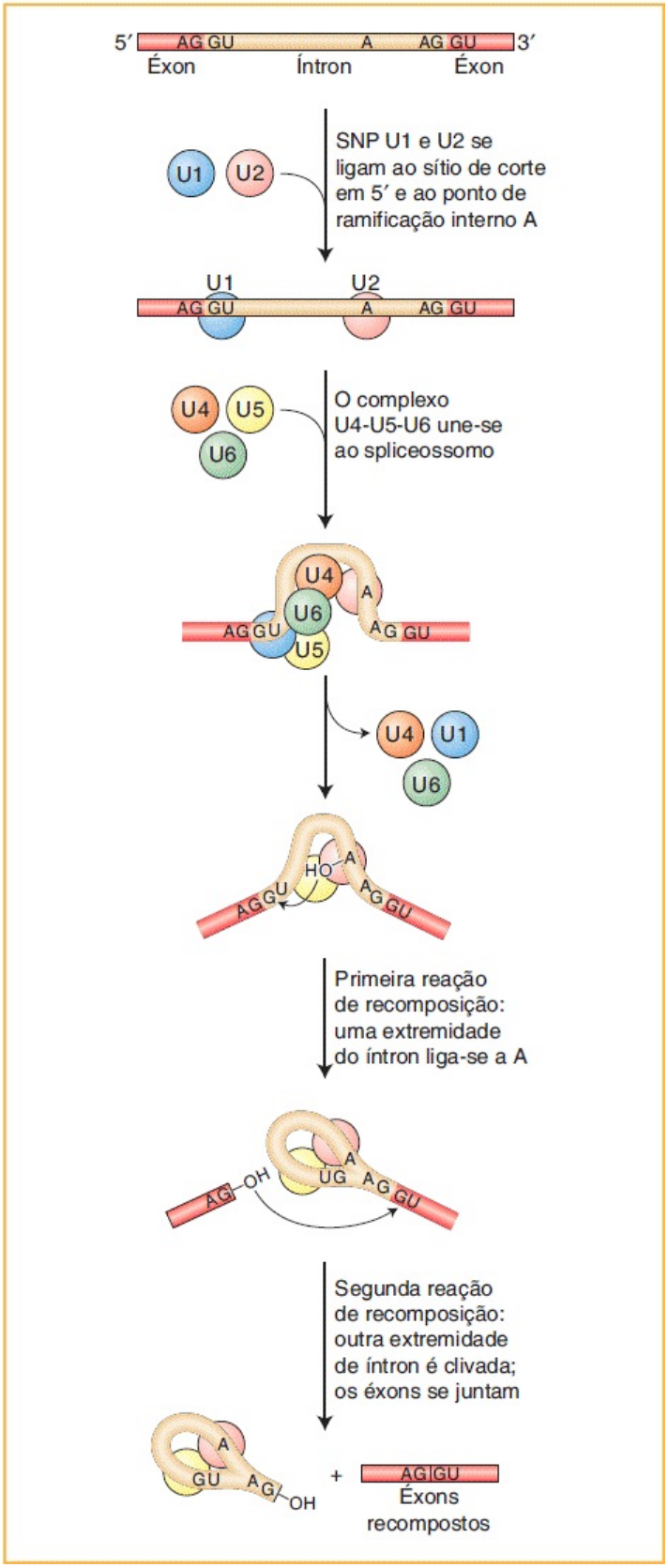


FIGURA 8.16 O spliceossomo é composto por diversos snRNP que se ligam sequencialmente ao RNA, geralmente nas posições conforme demonstradas. O alinhamento dos snRNP resulta das ligações de hidrogênio de suas moléculas de snRNA com as sequências complementares do íntron. Desse modo, os reagentes são adequadamente alinhados e as duas reações de recomposição podem ocorrer. A química destas reações pode ser observada em mais detalhes na Figura 8.17.

Reações na recomposição de éxons

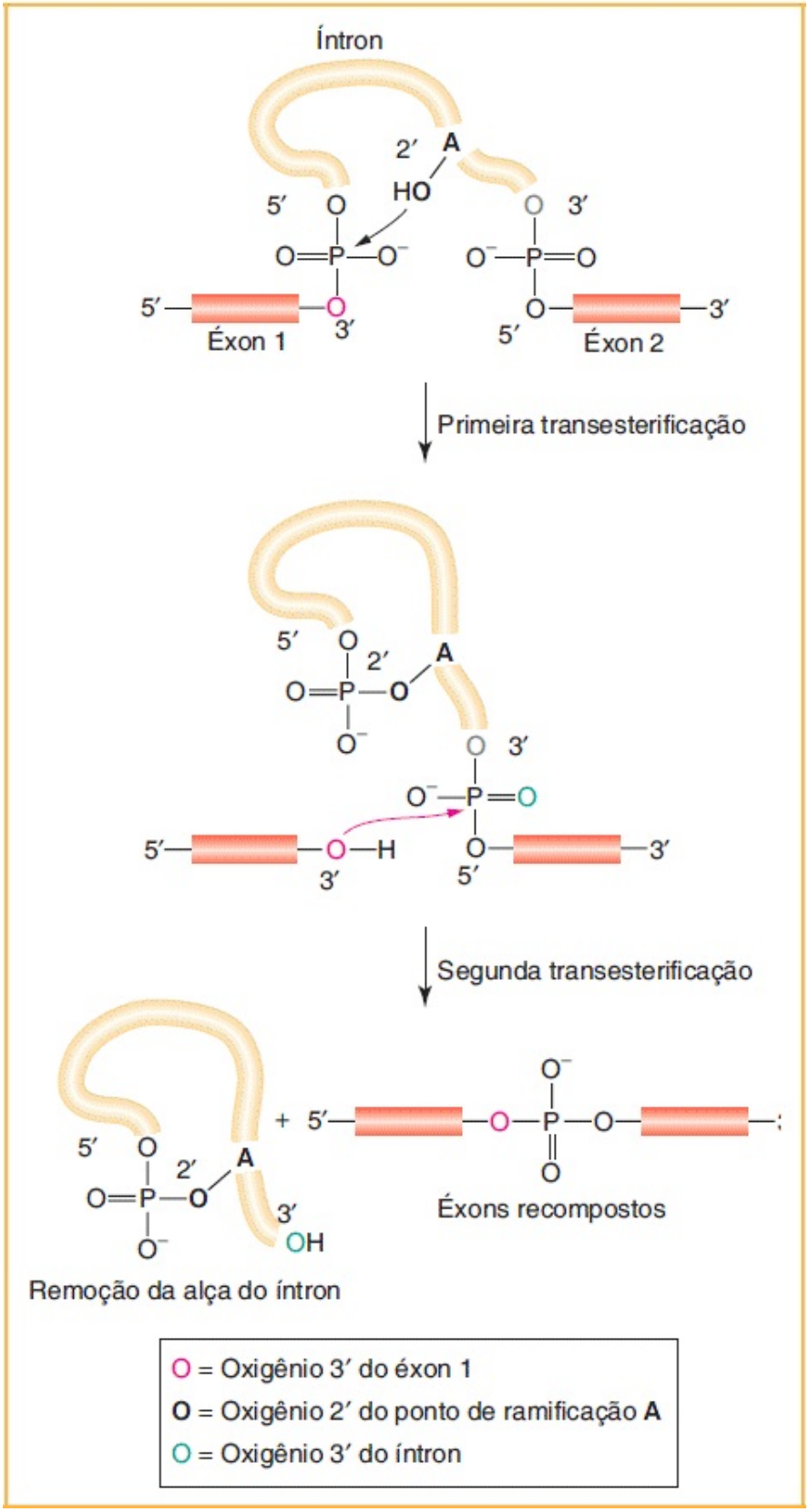


FIGURA 8.17 Duas reações de transesterificação ocorrem na recomposição do RNA: primeiramente, para unir a extremidade 5′ doadora do íntron ao ponto de ramificação interno (primeira reação na Figura 8.16) e, em segundo lugar, para unir os dois éxons (segunda reação na Figura 8.16).

Autorrecomposição de íntrons e o mundo do RNA

Dois casos excepcionais de processamento de RNA levaram a uma descoberta considerada por alguns tão importante quanto a da estrutura da dupla-hélice do DNA. Em 1981, Tom Cech e colaboradores relataram que, em um tubo de ensaio, o transcrito primário de um rRNA do protozoário ciliado *Tetrahymena* conseguiu remover um íntron de 413 nucleotídios *próprio seu* sem a adição de qualquer proteína (Figura 8.18). Subsequentemente, foi demonstrado que outros íntrons apresentam essa propriedade, e eles passaram a ser conhecidos como **íntrons de autorremoção**. Alguns anos antes, enquanto estudava o processamento do tRNA em bactérias, Sidney Altman identificou uma ribonucleoproteína (denominada RNase P) responsável pelo corte da molécula de pré-tRNA em um sítio específico. A grande surpresa surgiu quando eles determinaram que a atividade catalítica da RNase P estava localizada no componente RNA da enzima, não no componente proteico. Os achados de Cech e Altman são considerados marcantes, tendo em vista que assinalavam a primeira vez em que foi demonstrado que outras moléculas biológicas além de proteínas catalisam reações. Como tal, era justo que recebessem o Prêmio Nobel em Química em 1989.

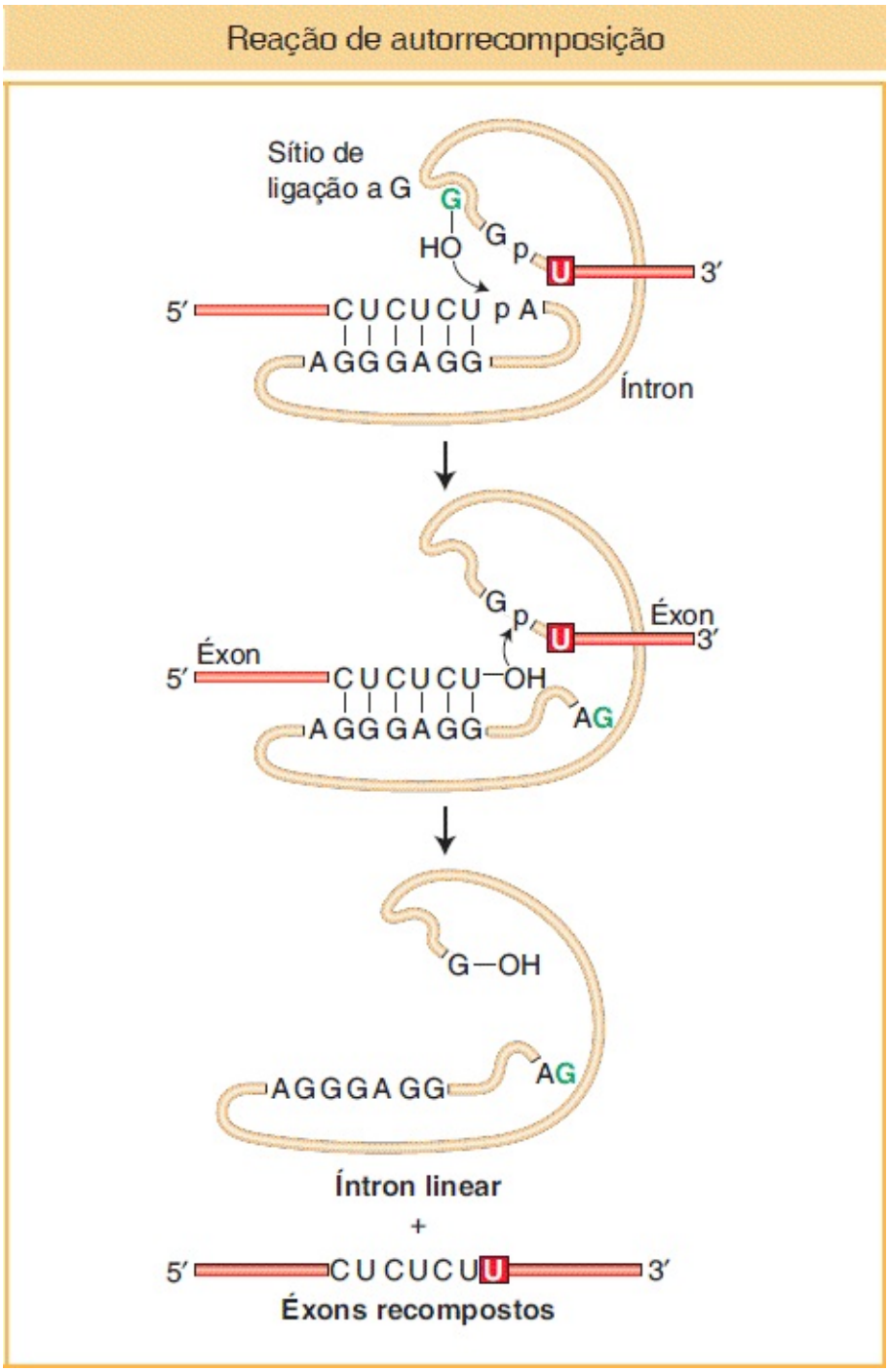


FIGURA 8.18 O íntron autorremovível do *Tetrahymena* executa duas reações de transesterificação para se autorremover do RNA.

A descoberta dos íntrons de autorremoção levou a um reexame do papel dos snRNA no spliceossomo. Os estudos mais recentes indicam que a remoção dos íntrons é catalisada pelos snRNA e não pelo componente proteico do

spliceossomo. Conforme veremos no Capítulo 9, os RNA no ribossomo (os rRNA), não as proteínas ribossômicas, apresentam o papel central na maior parte dos eventos importantes da síntese proteica. Os diversos exemplos de ribozimas forneceram evidências sólidas em relação a uma teoria denominada **mundo do RNA**, que sustenta que o RNA deve ter sido o material genético nas primeiras células, tendo em vista que apenas o RNA sabidamente codifica informação genética e catalisa reações biológicas.

> CONCEITO-CHAVE A remoção dos íntrons e a união dos éxons são catalisadas por moléculas de RNA. Em eucariotos, os snRNA do spliceossomo catalisam a remoção de íntrons do pré-mRNA. Alguns íntrons são autorremovíveis; nesses casos, o próprio íntron catalisa a sua própria remoção. RNA capazes de catálise são denominados ribozimas.

8.5 Pequenos RNA funcionais que regulam e protegem o genoma eucariótico

Em 2002, um dos principais periódicos científicos, a revista *Science*, nomeou o "Pequeno RNA" sua descoberta do ano. Os RNA aos quais eles estavam se referindo não eram os pequenos RNA descritos anteriormente, tais como os snRNA ou tRNA, que são considerados mantenedores e, como tal, são sintetizados constitutivamente. Em vez disso, esses outros pequenos RNA são sintetizados em resposta a alterações no estado do desenvolvimento de uma célula ou no ambiente celular. Atualmente sabemos que eles são criticamente importantes para a manutenção de um genoma estável e para a regulação da expressão gênica.

Os miRNA são importantes reguladores da expressão gênica

O primeiro desse tipo de pequenos RNA foi descoberto em 1993 por Victor Ambros e colegas enquanto estudavam o gene *lin-4* do nematódeo *C. elegans*.

Tendo em vista que mutações no *lin-4* resultavam em larvas anormais, foi formulada a hipótese de que esse gene codificava uma proteína que era necessária para o desenvolvimento normal da larva. Portanto, foi uma surpresa quando o grupo isolou o gene *lin-4* e relatou que, em vez de codificar uma proteína, ele produzia dois pequenos RNA de 22 nucleotídios e 61 nucleotídios. Em seguida, eles observaram que o RNA de 22 nucleotídios era produzido por meio do processamento do RNA maior de 61 nucleotídios. Finalmente, eles observaram que o RNA de 22 nucleotídios reprimia a expressão de determinados outros genes por meio do pareamento de bases com seus mRNA.

O produto do RNA de 22 nucleotídios do gene *lin-4* foi o primeiro membro a ser descoberto de uma classe muito grande de RNA, denominados microRNA (miRNA), agora sabidamente presentes nos genomas de plantas e de animais. A maior parte dos miRNA atua para reprimir a expressão de genes. De fato, estima-se que cada um dos genomas de plantas e de animais apresente até 1.000 miRNA que, por sua vez, regulam a expressão de milhares de genes. Assim como o produto de *lin-4*, muitos miRNA são inicialmente transcritos pela RNA polimerase II como um RNA mais longo de um gene que produz apenas um produto de RNA. O RNA mais longo assume uma estrutura em alça com haste de filamento duplo com uma base errada na haste (Figura 8.19). O RNA é processado no núcleo até uma forma menor, porém ainda não final, e em seguida é exportado para o citoplasma. Ali, duas máquinas biológicas, ambas com a capacidade de clivar o RNA, participam em um processo de duas etapas. Uma máquina, denominada **Dicer**, reconhece as moléculas de **RNA de filamento duplo (dsRNA)** e as cliva em produtos de aproximadamente 22 nucleotídios. Uma segunda máquina, denominada **RISC** (**complexo de silenciamento induzido por RNA**; do inglês, *RNA-induced silencing complex*), liga-se ao dsRNA curto e o desenrola formando o miRNA de filamento único biologicamente ativo. O miRNA, ainda ligado ao RISC, liga-se a mRNA complementares. O RISC em seguida reprime a tradução desses mRNA em proteínas ou remove a cauda poli(A), o que acelera a degradação do mRNA. No exemplo demonstrado na Figura 8.19, o miRNA do *lin-4* se liga aos mRNA de *lin-14* e *lin-28* e reprime a sua tradução.

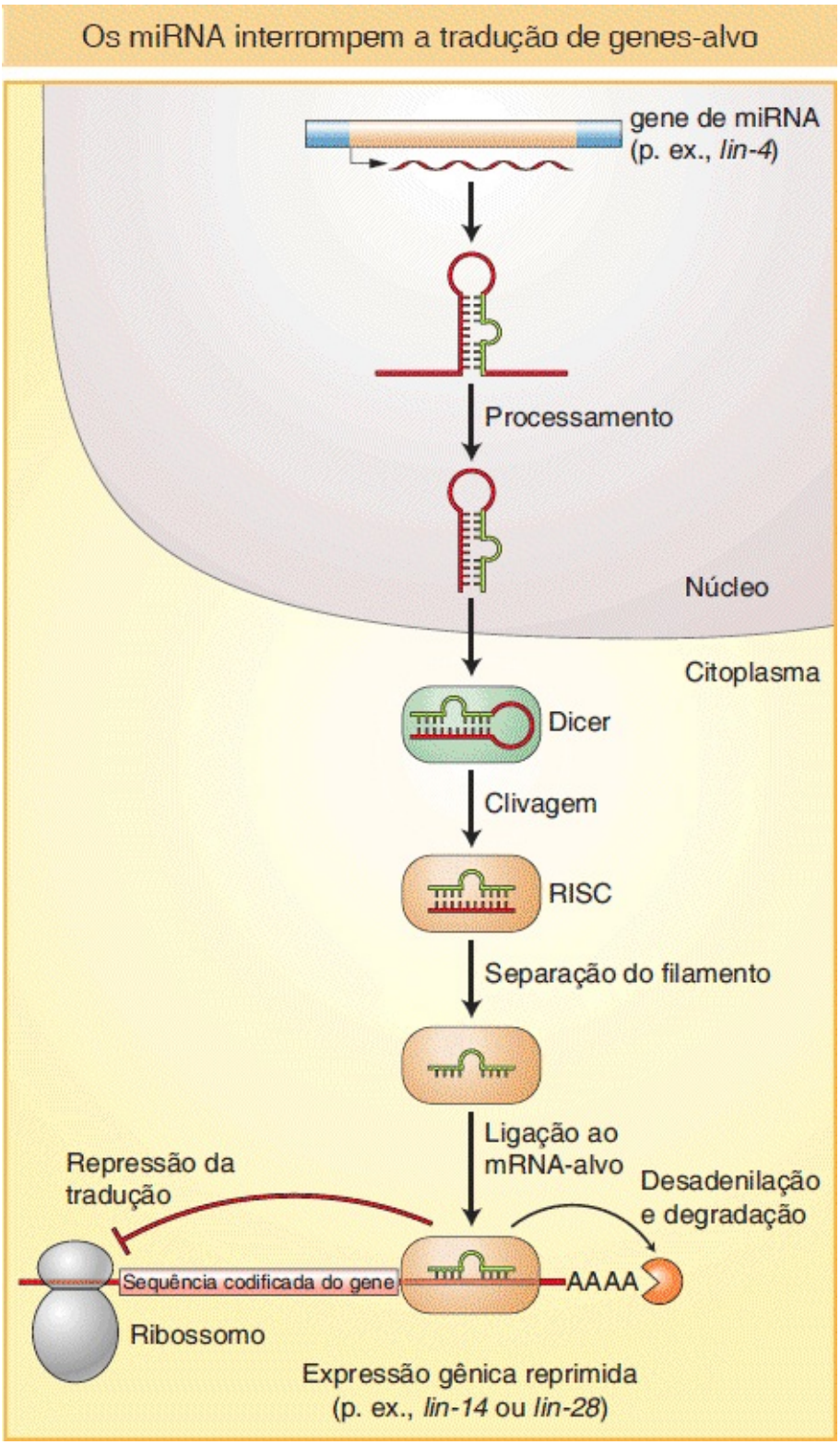


FIGURA 8.19 Os miRNA são sintetizados pela RNA pol II como RNA mais longos, que são processados em diversas etapas até a sua forma madura. Uma vez totalmente processados, os miRNA se ligam ao RISC e direcionam as suas atividades para reduzir a expressão de mRNA complementares por meio da repressão de

sua tradução ou da promoção de sua degradação.

Você aprenderá mais a respeito da função dos miRNA nos Capítulos 12 e 13. O ponto-chave a ser relembrado é que parte de um miRNA é complementar ao RNA do gene que ele regula. Quando o gene regulado precisa ser desligado ou a sua expressão precisa ser reduzida, o gene do miRNA é transcrito em RNA e aquele RNA se liga ao RNA do gene regulado, interferindo na tradução em proteína ou promovendo a sua degradação.

> CONCEITO-CHAVE Os miRNA são processados a partir de transcritos mais longos da RNA pol II pela Dicer, que se liga aos RNA de filamento duplo. O miRNA de filamento único biologicamente ativo se liga ao RISC e o direciona até sequências complementares em mRNA codificadores de proteínas, onde o RISC reprime a tradução ou promove a degradação do mRNA.

Os siRNA asseguram a estabilidade do genoma

Os cientistas logo encontraram um caso diferente de dsRNA que poderia reprimir a expressão gênica antes da tradução. Esse achado levou à descoberta de um segundo tipo de RNA curto, os siRNA. Esse segundo tipo de RNA curto apresenta origem e função muito diferentes dos miRNA. Contrariamente aos miRNA, um siRNA silencia o gene que o produz. Portanto, ele não é utilizado para regular outros genes, mas, em vez disso, para desativar elementos genéticos indesejáveis que se inserem no genoma. Os referidos elementos indesejáveis podem ser genes em um vírus infectante, ou podem ser elementos genéticos internos denominados transpósons, a respeito dos quais você aprenderá no Capítulo 15.

Em 1998, 5 anos após a descoberta dos miRNA, Andrew Fire e Craig Mello relataram que haviam encontrado um modo potente de desligar seletivamente genes, também no nematódeo *C. elegans*. Fire e Mello descobriram que, ao injetar cópias do dsRNA de um gene de *C. elegans* em embriões de *C. elegans*, eles conseguiram bloquear a síntese do produto proteico daquele gene (Figura

8.20 A). O desligamento seletivo do gene por meio desse procedimento é denominado **silenciamento gênico**. O dsRNA havia sido sintetizado no laboratório e era composto por um filamento de RNA senso (codificador) e um **filamento de RNA antissenso** complementar. Em seu experimento inicial, Fire e Mello injetaram cópias de dsRNA do gene *unc-22* em embriões de *C. elegans* e observaram, na medida em que os embriões cresciam até adultos, que se contraíam e que apresentavam defeitos musculares. Esse resultado era animador, tendo em vista que sabidamente o *unc-22* codificava uma proteína muscular e mutantes nulos de *unc-22* demonstravam as mesmas contrações e os mesmos defeitos musculares. Consideradas em conjunto, essas observações indicaram que o dsRNA injetado evitou a produção da proteína Unc-22. Em virtude de sua descoberta sobre um novo modo de silenciar genes, Fire e Mello receberam o Prêmio Nobel de Fisiologia ou Medicina em 2004.

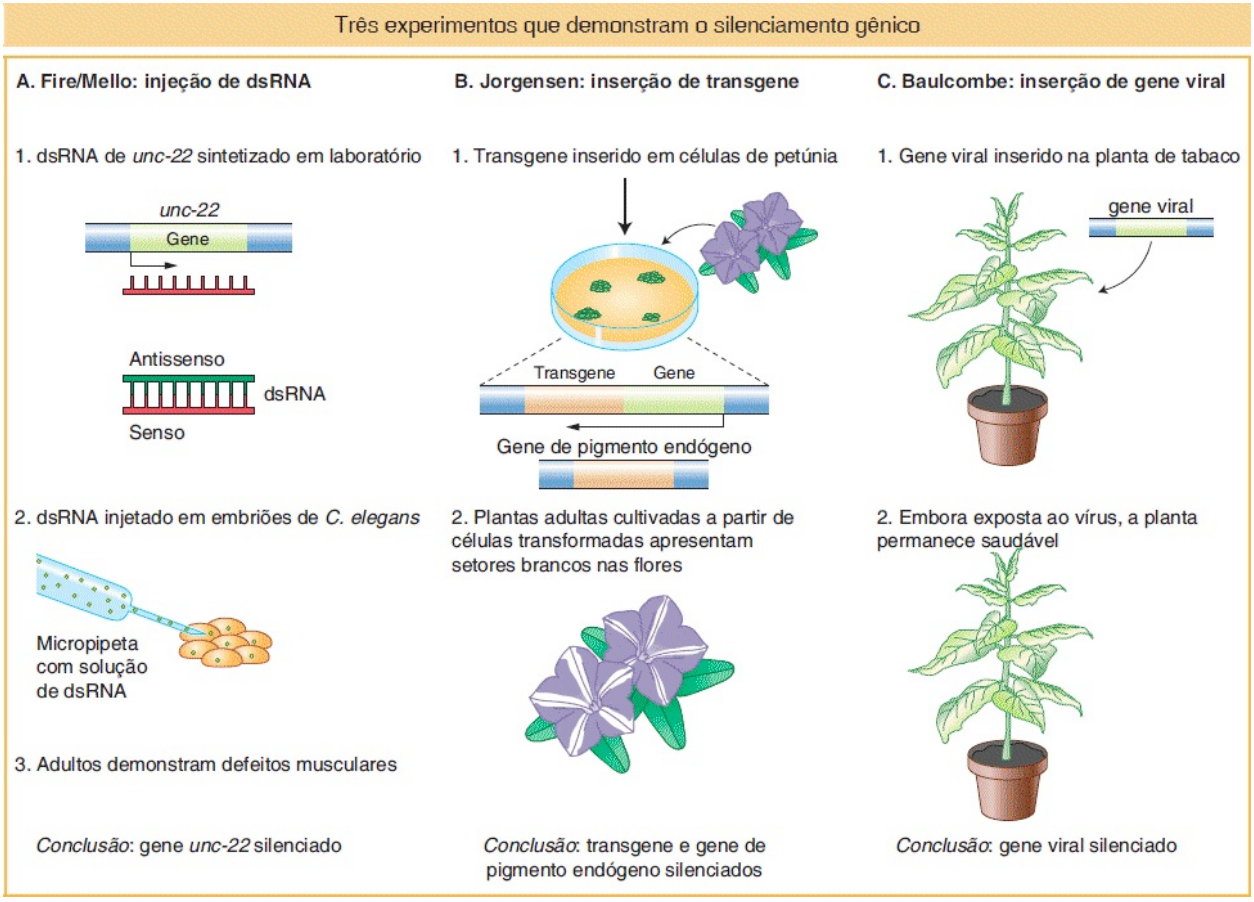


FIGURA 8.20 Três experimentos revelam características fundamentais do silenciamento gênico. **A.** Fire e

Mello demonstraram que cópias de dsRNA podem silenciar seletivamente genes em *C. elegans*. **B.** Jorgensen descobriu que um transgene pode silenciar um gene endógeno de petúnia necessário para a cor das flores. **C.** Baulcombe demonstrou que plantas com uma cópia de um transgene viral eram resistentes à infecção viral e produziam siRNA complementares ao genoma viral.

Em vez de uma cópia de dsRNA de um gene, o que aconteceria se uma cópia de DNA de um gene normalmente observado em um organismo fosse inserida no seu genoma? Em um referido experimento, o gene introduzido seria um exemplo de um **transgene**, que é uma abreviação de "gene transformado". Um transgene é um gene que foi introduzido nos cromossomos de um organismo em laboratório. Um organismo que contém um transgene em seu genoma é denominado organismo **transgênico** ou **organismo geneticamente modificado** (ou o termo abreviado popular **OGM**). Esse experimento foi realmente realizado em 1990 por Rich Jorgensen, um cientista de plantas que estudava a cor das flores em petúnias.

Um dos maiores prazeres de realizar pesquisas científicas é a observação de um resultado completamente inesperado. Isso é precisamente o que ocorreu com Jorgensen após ter inserido um gene de petúnia que codifica uma enzima necessária para a síntese do pigmento floral roxo-azulado em uma planta de petúnia normal que apresenta flores roxo-azuladas (Figura 8.20 B). Ele esperava que a cor das flores dessa planta transgênica não fosse alterada. Afinal, a planta transgênica apresentava dois genes necessários para a produção do pigmento — um em seu *locus* habitual no genoma da petúnia (denominado **gene endógeno**; na Figura 8.20 B ele é denominado gene de pigmento), mais o transgene introduzido, que foi inserido em outro local no genoma. Entretanto, em vez de flores roxas, as plantas transgênicas demonstraram os padrões florais incomuns demonstrados na Figura 8.21. Em um desfecho totalmente inesperado, o transgene acionou a supressão de ambos o transgene e o gene de pigmento endógeno, resultando em flores brancas ou, mais comumente, setores florais brancos. Esse fenômeno é denominado **cossupressão**, tendo em vista que a expressão de ambos o transgene introduzido e a cópia endógena é suprimida.

Para uma revisão, a introdução de uma cópia de dsRNA de um gene ou do próprio gene dentro de um organismo pode silenciar aquele gene. Para compreender o motivo pelo qual esses diferentes experimentos levaram ao mesmo

resultado, cientistas formularam a hipótese de que a inserção do transgene levou à síntese de RNA antissenso, que poderia se complementar com um RNA senso para produzir dsRNA. Tendo em vista que os cientistas não podem controlar onde os transgenes são inseridos, alguns transgenes acabarão próximos a genes em uma orientação oposta (Figura 8.22). A transcrição iniciada no promotor gênico pode prosseguir a "leitura" através do transgene e produzir um RNA "quimérico" muito longo, que contém tanto o filamento com sentido (senso) do gene quanto o filamento antissenso do transgene. RNA bifilamentar em seguida será formado quando a parte antissenso do RNA longo hibridizar com o RNA senso produzido pelo transgene ou pelo gene endógeno.

Portanto, o dsRNA é uma característica comum desse tipo se silenciamento gênico. Entretanto, a função desse processo claramente não é desligar genes introduzidos pelos cientistas. Qual é o papel normal desse tipo de silenciamento gênico na célula? Uma indicação importante surgiu a partir dos experimentos conduzidos por outro cientista de plantas, David Baulcombe, que estava investigando o motivo pelo qual as plantas de tabaco que eram modificadas para expressar um gene viral eram resistentes à infecção subsequente pelo vírus. Nesse experimento, o gene viral é outro exemplo de um transgene, o qual, nesse caso, foi introduzido no genoma do tabaco (ver Figura 8.20 C). Uma diferença-chave entre este e o experimento com petúnias foi que as plantas do tabaco normalmente não apresentam um gene viral no seu genoma. Assim, esse experimento sugeriu que esse tipo de silenciamento gênico atue para silenciar vírus invasores.

Flores de petúnia demonstrando cossupressão

FIGURA 8.21 A. O fenótipo do tipo selvagem (nenhum transgene). **B** e **C.** Os assim denominados fenótipos de cossupressão que resultam da transformação da petúnia do tipo selvagem demonstrada na parte **A** com um gene de petúnia necessário para a pigmentação. Nas regiões incolores, ambos o transgene e a cópia cromossômica do mesmo gene foram inativados. (Richard Jorgensen, Department of Plant Biology, Carnegie Institution for Science.)

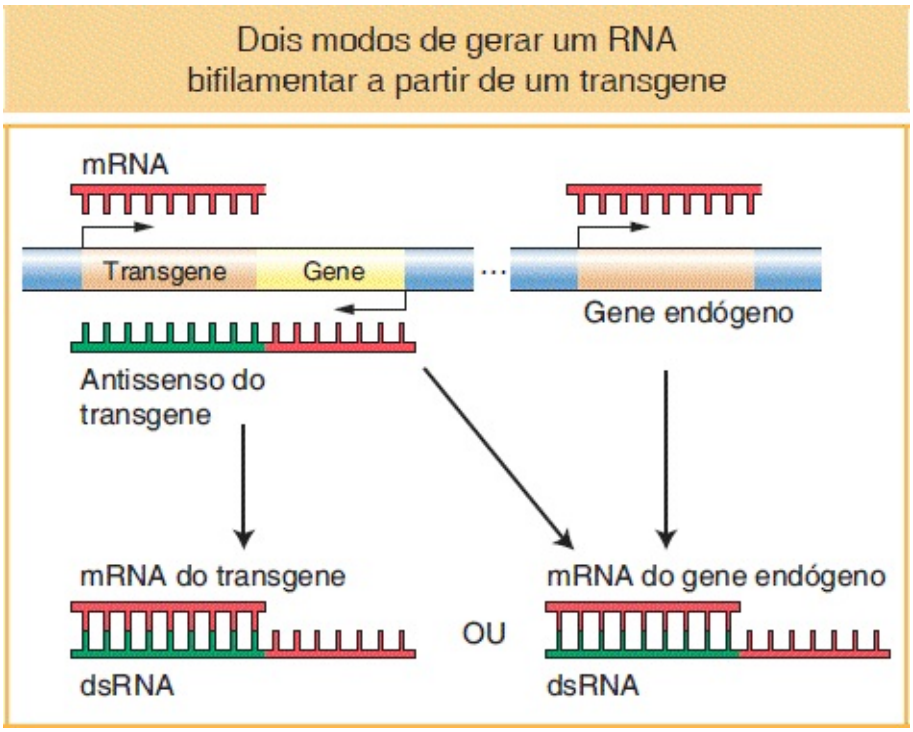


FIGURA 8.22 A inserção de um transgene pode levar à produção de RNA bifilamentar (dsRNA) se o transgene for inserido na extremidade de um gene em orientação oposta. O RNA antissenso produzido quando o gene vizinho é transcrito pode se ligar ao mRNA do próprio transgene ou do gene endógeno para produzir dsRNA.

Baulcombe e seus colaboradores observaram que as plantas resistentes, e apenas as plantas resistentes, produziam grandes quantidades de pequenos RNA, de 25 nucleotídios de comprimento, que eram complementares ao genoma viral. Significativamente, também foi observado que os pequenos RNA relacionados com os genes endógenos estão presentes durante o silenciamento gênico no nematódeo e na petúnia. Os RNA curtos gerados durante a resistência viral e o silenciamento gênico associados aos dsRNA ou transgenes injetados atualmente são denominados coletivamente **pequenos RNA de interferência** (**siRNA**; do inglês, *small interfering RNAs*). O fenômeno que resulta no silenciamento gênico

e na resistência viral por meio da produção de siRNA é denominado **RNA de interferência** (**RNAi**; do inglês, *RNA interference*). Os RNA curtos (21 a 31 nucleotídios de comprimento) atualmente são classificados como um de três tipos, dependendo da sua biogênese: miRNA ou siRNA (ambos com 21 a 25 nucleotídios), ou os recentemente descobertos **RNA de interação *piwi*** (**piRNA** [do inglês, *piwi-interacting* RNAs], 24 a 31 nucleotídios). Tendo em vista que o mecanismo da síntese de piRNA ainda está em investigação, enfocaremos os mais bem-caracterizados miRNA e siRNA.

Mecanismos semelhantes geram siRNA e miRNA

Conforme vimos, os siRNA podem ter origem a partir de uma cópia antissenso de qualquer fonte de mRNA no genoma: desde os genes endógenos até os transgenes e até os vírus invasores. Entretanto, a fonte mais provável de RNA antissenso não são os genes do próprio organismo, mas, em vez disso, um DNA estranho que se insere no genoma. Nesse sentido, seria correto pensar nos siRNA como o produto de um sistema imune do genoma que detecta a inserção de DNA estranho, em alguns casos, promove a síntese de mRNA antissenso. A complementaridade entre os RNA senso e antissenso produz dsRNA, os quais, assim como na via de miRNA, são reconhecidos pela Dicer e clivados em produtos bifilamentares curtos que são ligados pelo RISC (Figura 8.23). Assim como com os miRNA, RISC desenrola o produto em um siRNA unifilamentar biologicamente ativo que direciona RISC para mRNA complementares, de modo que eles possam ser degradados. Contrariamente aos miRNA, a complementaridade entre os siRNA e os mRNA é perfeita; não existem incompatibilidades. Isso ocorre em virtude de suas origens diferentes: siRNA são derivados do mesmo gene, enquanto miRNA advêm de um gene diferente. Essa diferença provavelmente é responsável pelos desfechos diferentes: miRNA direcionam RISC para reprimir a tradução de um mRNA ou degradar mRNA quando eles estão sendo traduzidos, enquanto siRNA direcionam RISC para degradar o mRNA diretamente.

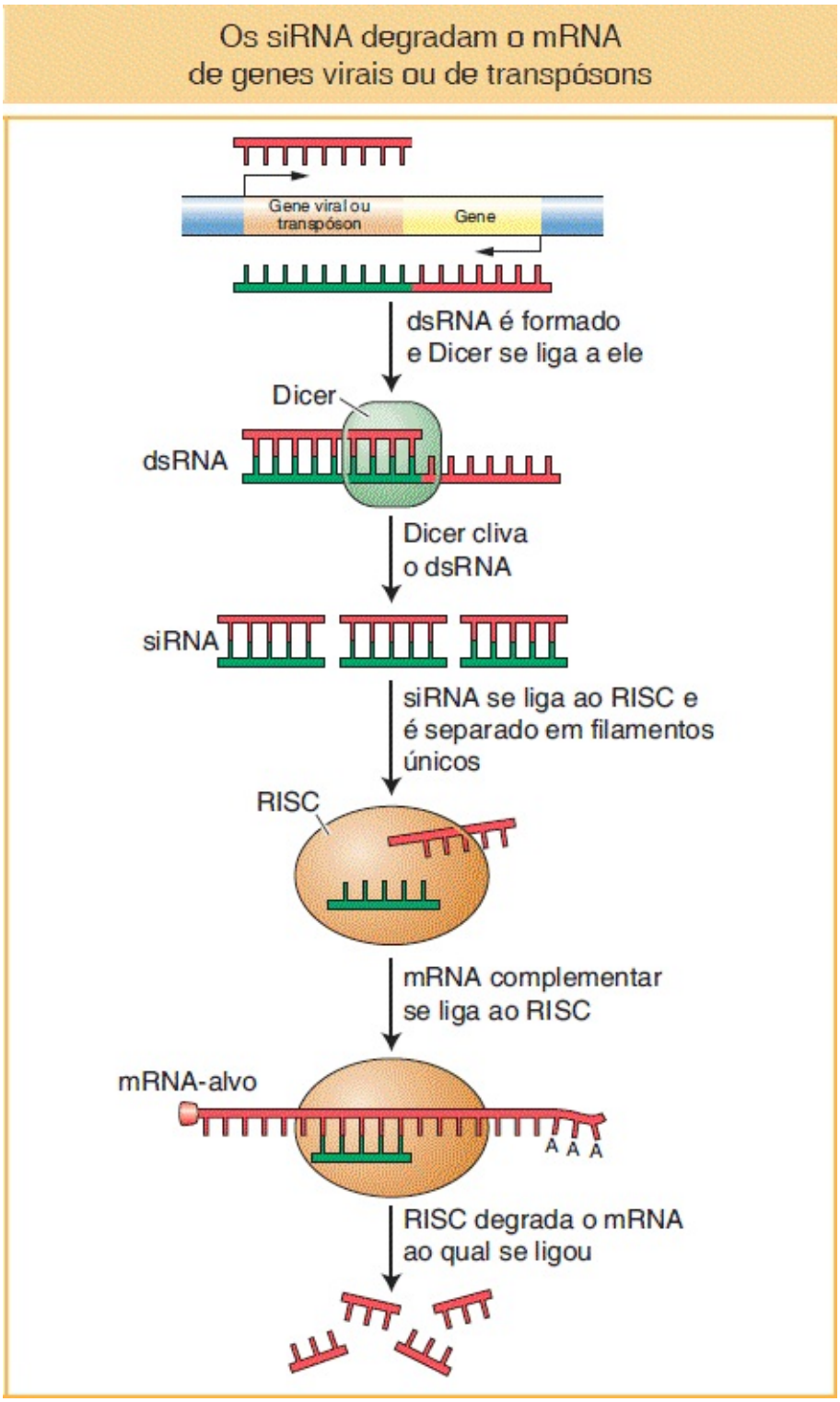


FIGURA 8.23 Na via de RNA de interferência, o RNA bifilamentar (dsRNA) interage especificamente com o complexo Dicer, que corta o dsRNA. O complexo de silenciamento induzido por RNA (RISC) utiliza os pequenos dsRNA para encontrar e destruir mRNA homólogo transcrito a partir do DNA-alvo, reprimindo, assim, a expressão gênica.

Conforme discutido anteriormente, a produção de siRNA provavelmente desempenha um papel importante na defesa viral. Entretanto, seu papel mais importante pode ser proteger o material hereditário de um organismo contra elementos genéticos no seu próprio genoma. No Capítulo 15, você aprenderá a respeito dos elementos de transposição que constituem uma enorme fração dos genomas de eucariotos multicelulares, incluindo seres humanos. Esses elementos podem amplificar a si próprios e se mover para locais novos, criando uma óbvia ameaça à integridade do genoma. Assim como a introdução dos transgenes pelos cientistas, a movimentação dos elementos de transposição para novos locais cromossômicos pode acionar a produção de siRNA por meio da geração de dsRNA. Os siRNA finalmente inativam os elementos de transposição em parte ao prevenir a síntese dos produtos proteicos necessários para a sua movimentação e amplificação.

> CONCEITO-CHAVE O RNA antissenso com frequência é formado em resposta à inserção de DNA estranho no genoma. Dicer detecta o RNA bifilamentar que se forma entre o RNA antissenso e senso e o processa em RNA curtos. RISC se liga ao RNA curto e o desenrola para formar siRNA biologicamente ativo. O siRNA direciona RISC para um mRNA perfeitamente complementar, que é degradado, silenciando, assim, a expressão do DNA estranho.

RESUMO

Sabemos que a informação não é transferida diretamente do DNA para a proteína, tendo em vista que, em uma célula eucariótica, o DNA está no núcleo, enquanto a proteína é sintetizada no citoplasma. A transferência da informação do DNA para a proteína necessita de um intermediário. O intermediário é o RNA.

Embora o DNA e o RNA sejam ácidos nucleicos, o RNA difere do DNA no sentido em que (1) normalmente é um filamento único, em vez de uma dupla-hélice, (2) seus nucleotídios contêm o açúcar ribose em vez de desoxirribose, (3)

ele apresenta a base pirimidínica uracila, em vez de timina, e (4) pode atuar como um catalisador biológico.

A semelhança do RNA com o DNA sugere que o fluxo de informação do DNA para o RNA depende da complementaridade das bases, o que também é a chave para a replicação do DNA. Um filamento-molde de DNA é copiado, ou transcrito, em um RNA funcional (tal como RNA transportador ou RNA ribossômico), que nunca é traduzido em polipeptídios, ou em um RNA mensageiro, a partir do qual as proteínas são sintetizadas.

Em procariotos, todas as classes de RNA são transcritas por uma única RNA polimerase. Essa enzima multissubunitária inicia a transcrição por meio da ligação ao DNA em promotores que contêm sequências específicas em —35 e —10 bases antes do sítio de início da transcrição em +1. Após a sua ligação, a RNA polimerase desenrola localmente o DNA e inicia a incorporação de ribonucleotídios que são complementares ao filamento-molde de DNA. A cadeia cresce no sentido 5′ para 3′ até que um de dois mecanismos, intrínseco ou rho-dependente, leve à dissociação da polimerase e do RNA do molde de DNA. Conforme veremos no Capítulo 9, na ausência de um núcleo, os RNA procarióticos que codificam proteínas são traduzidos enquanto estão sendo transcritos.

Em eucariotos, existem três RNA polimerases diferentes; apenas a RNA polimerase II transcreve mRNA. Em geral, as fases de iniciação, alongamento e término da síntese de RNA em eucariotos se assemelham às dos procariotos. Entretanto, existem diferenças importantes. A RNA polimerase II não se liga diretamente ao promotor no DNA, mas, em vez disso, aos GTF, um dos quais reconhece a sequência TATA na maior parte dos promotores eucarióticos. A RNA polimerase II é uma molécula muito maior que a sua correspondente procariótica. Ela contém diversas subunidades que atuam não apenas para alongar o transcrito primário de RNA, mas também para coordenar os eventos de processamento extensivos que são necessários para produzir o mRNA maduro. Esses eventos de processamento são a adição do *cap* em 5′, a remoção de íntrons e a união de éxons pelos spliceossomos, e a clivagem em 3′ seguida por poliadenilação. Parte do centro da RNA polimerase II, o domínio carboxi-terminal (CTD), é

posicionado idealmente para interagir com o RNA nascente na medida em que ele surge a partir da polimerase. Por meio do CTD, a RNA polimerase II coordena os diversos eventos de síntese e de processamento do RNA.

As descobertas dos últimos 20 anos revelaram a importância de novas classes de RNA funcionais. O RNA, o qual se chegou a acreditar que fosse um mensageiro modesto, atualmente é reconhecido como um participante versátil e dinâmico em muitos processos celulares. A descoberta de íntrons autorremovíveis demonstrou que o RNA pode atuar como um catalisador, de modo muito semelhante às proteínas. Desde a descoberta dessas ribozimas, a comunidade científica começou a prestar mais atenção ao RNA. Pequenos RNA nucleares, os RNA não codificadores no spliceossomo, atualmente são reconhecidos como promotores da atividade catalítica para remover íntrons e unir os éxons. O século 20 terminou com a descoberta de que duas outras classes de RNA funcional, miRNA e siRNA, associam-se aos complexos de silenciamento induzidos por RNA (RISC) e têm por alvo o mRNA celular complementar para a repressão (no caso do miRNA) ou para a destruição (no caso do siRNA).

TERMOS-CHAVE

alongamento
bolha de transcrição
cap (revestimento)
cauda poli(A)
complexo de pré-iniciação (PIC)
constitutiva
cossupressão
Dicer
domínio carboxi-terminal
downstream
éxon
experimento de pulso-caça
expressão gênica

fator geral de transcrição (GTF)
fator sigma (σ)
filamento codificador
filamento de RNA antissenso
gene endógeno
holoenzima RNA polimerase
iniciação
íntron
íntron de autorremoção
microRNA (miRNA)
molde
mundo do RNA
organismo geneticamente modificado (transgênico) (OGM)
pequenos RNA de interferência (siRNA)
pequenos RNA nucleares (snRNA)
processamento cotranscricional
processamento pós-transcricional
processamento do RNA
promotor
proteína de ligação a TATA (TBP)
proteoma
recomposição alternativa
recomposição ou *splicing*
recomposição ou *splicing* do RNA
região 3′ não traduzida (3′ UTR)
região 5′ não traduzida (5′ UTR)
regra GU–AG
ribose
ribozima
RISC (complexo de silenciamento induzido por RNA)
RNA de filamento duplo (bifilamentar) (dsRNA)
RNA de interação *piwi* (piRNA)

RNA de interferência (RNAi)
RNA funcional
RNA mensageiro (mRNA)
RNA não codificador de proteínas (ncRNA)
RNA não codificadores longos (lncRNA)
RNA polimerase
RNA ribossômico (rRNA)
RNA transportador (tRNA)
sequência consenso
silenciamento gênico
spliceossomo
TATA boxe
término
transcrição
transcrito
transcrito primário (pré-mRNA)
transgene
upstram
uracila (U)

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 8.3, por que as setas em relação aos genes 1 e 2 estão apontando em sentidos opostos?
2. Na Figura 8.5, desenhe o "gene um" em resolução muito superior com os componentes a seguir: DNA, RNA polimerase(s), RNA(s).
3. Na Figura 8.6, descreva onde o promotor gênico está localizado.
4. Na Figura 8.9 B, escreva uma sequência que possa formar a estrutura de alça em grampo.
5. Como você sabe que os eventos na Figura 8.13 estão ocorrendo no núcleo?

6. Na Figura 8.15, qual você acredita que seja o efeito de uma mutação de G para A no primeiro resíduo G do íntron?
7. Ao comparar as Figuras 8.16 e 8.17, avalie qual é/quais são a(s) função(ões) das proteínas U1 a U6.
8. Ao comparar as Figuras 8.16 e 8.17 com a Figura 8.18, especule quais características do RNA possibilitam a autorrecomposição (ou seja, na ausência de proteínas).
9. Na Figura 8.20, estão demonstradas três situações muito diferentes, todas resultando em silenciamento gênico. O que essas situações apresentam em comum para tornar isso possível?
10. Na Figura 8.22, demonstre como o RNA bifilamentar é capaz de silenciar o transgene. O que deve ocorrer para que o transgene também silencie o gene celular flanqueador (em amarelo)?

PROBLEMAS BÁSICOS

11. Em procariotos e eucariotos, descreva o que mais está ocorrendo com o RNA enquanto a RNA polimerase está sintetizando um transcrito a partir do molde de DNA.
12. Liste três exemplos de proteínas que atuam nos ácidos nucleicos durante a transcrição.
13. Qual é a função primária do fator sigma? Existe, em eucariotos, uma proteína análoga ao fator sigma?
14. Você identificou uma mutação em uma levedura, um eucarioto unicelular, que previne o revestimento da extremidade 5′ do transcrito de RNA. Entretanto, para a sua grande surpresa, todas as enzimas necessárias para o revestimento são normais. Você determina que a mutação é, em vez disso, em uma das subunidades da RNA polimerase II. Qual subunidade é mutante, e como essa mutação resulta na falha em adicionar um *cap* ao RNA de levedura?
15. Por que o RNA é produzido apenas a partir do filamento-molde de DNA e

não a partir de ambos os filamentos?

16. Um plasmídio linear contém apenas dois genes, os quais são transcritos em sentidos opostos, cada um partir da extremidade e em direção ao centro do plasmídio. Desenhe diagramas:

a. Do DNA do plasmídio, demonstrando as extremidades 5′ e 3′ dos filamentos de nucleotídios.

b. Do filamento-molde em relação a cada gene.

c. Das posições dos sítios de início da transcrição.

d. Dos transcritos, demonstrando as extremidades 5′ e 3′.

17. Existem semelhanças entre as bolhas de replicação do DNA e as bolhas de transcrição observadas em eucariotos? Explique.

18. Quais das declarações a seguir são verdadeiras a respeito do mRNA eucariótico?

a. O fator sigma é essencial para a correta iniciação da transcrição.

b. O processamento do mRNA nascente pode ter início antes de sua transcrição ser concluída.

c. O processamento ocorre no citoplasma.

d. O término é realizado por meio da pela utilização de uma alça em grampo ou por meio da utilização do fator rhô.

e. Muitos RNA podem ser transcritos simultaneamente a partir de um molde de DNA.

19. Uma pesquisadora estava induzindo mutações em células procarióticas por meio da inserção de segmentos de DNA. Desse modo, ela realizou a mutação a seguir:

Original	TTGACAT <u>15 para 17 pb</u> TATAAT
Mutante	TATAAT <u>15 para 17 pb</u> TTGACAT

a. O que essa sequência representa?

b. Qual você prevê que seja o efeito de uma tal mutação? Explique.

20. Você aprenderá mais a respeito da engenharia genética no Capítulo 10, mas, por enquanto, coloque seu capacete de engenheiro genético e tente resolver este problema. *E. coli* é amplamente utilizada em laboratórios para produzir proteínas de outros organismos.

a. Você isolou um gene de levedura que codifica uma enzima metabólica e deseja produzir essa enzima em *E. coli*. Você suspeita que o promotor de levedura não funcionará em *E. coli*. Por quê?

b. Após substituir o promotor de levedura por um promotor de *E. coli*, você com prazer detecta o RNA do gene de levedura, mas está confuso, porque o RNA apresenta quase o dobro do comprimento do mRNA desse gene isolado de levedura. Explique por que esse resultado pode ter ocorrido.

21. Desenhe um gene procariótico e seu produto RNA. Assegure-se de incluir o promotor, o sítio de início da transcrição, o sítio de término da transcrição, as regiões não traduzidas e as extremidades 5′ e 3′ marcadas.

22. Desenhe um gene eucariótico com dois íntrons e seus produtos de pré-mRNA e mRNA. Assegure-se de incluir todas as características do gene procariótico incluídas em sua resposta ao Problema 19, mais os eventos de processamento necessários para produzir o mRNA.

23. Um determinado gene codificador de proteína de *Drosophila* apresenta um íntron. Se uma amostra grande de alelos nulos desse gene for examinada, alguns sítios mutantes serão esperados:

a. Nos éxons?

b. No íntron?

c. No promotor?

d. No limite íntron–éxon?

24. O que são íntrons de autorremoção e por que sua existência ampara a teoria de que o RNA surgiu antes da proteína?

25. Antibióticos são fármacos que matam seletivamente bactérias sem prejudicar os animais. Muitos antibióticos atuam ao se ligar seletivamente a determinadas proteínas que são críticas para a função bacteriana. Explique por que alguns dos antibióticos de mais sucesso têm por alvo a RNA

polimerase bacteriana.

26. Descreva quatro tipos de RNA que realizam funções distintas.

PROBLEMAS DESAFIADORES

27. Os dados a seguir representam as composições de bases do DNA bifilamentar de duas espécies bacterianas diferentes e seus produtos de RNA obtidos em experimentos conduzidos *in vitro*:

Espécie	$\frac{(A + T)}{(G + C)}$	$\frac{(A + U)}{(G + C)}$	$\frac{(A + G)}{(U + C)}$
Bacillus subtilis	1,36	1,30	1,02
E. coli	1,00	0,98	0,80

a. A partir desses dados, determine se o RNA dessas espécies é copiado a partir de um filamento único ou de ambos os filamentos do DNA. Desenhe um diagrama para demonstrar como você resolve este problema.

b. Como você poderia dizer se o próprio RNA é unifilamentar ou bifilamentar?

28. Um gene humano foi inicialmente identificado como apresentando três éxons e dois íntrons. Os éxons são de 456, 224 e 524 pb, enquanto os íntrons são de 2,3 kb e 4,6 kb.

a. Desenhe esse gene, demonstrando o promotor, os íntrons, os éxons e os sítios de início e término da transcrição.

b. Surpreendentemente, observa-se que esse gene codifica não um, mas dois mRNA que apresentam apenas 224 nucleotídios em comum. O mRNA original é de 1.204 nucleotídios, e o novo mRNA tem 2.524 nucleotídios. Utilize o seu desenho para demonstrar como essa região única do DNA

pode codificar esses dois transcritos.

29. Enquanto trabalha em seu laboratório, você isola um mRNA de *C. elegans* que você suspeita ser essencial para que os embriões se desenvolvam com sucesso. Com a presunção de que você é capaz de tornar o mRNA um RNA bifilamentar, desenhe um experimento para testar a sua hipótese.

30. O glifosato é um herbicida utilizado para matar ervas daninhas. Ele é o principal componente de um produto fabricado pela Monsanto Company denominado Roundup®. O glifosato mata as plantas por meio da inibição de uma enzima da via shikimate denominada EPSPS. Esse herbicida é considerado seguro, tendo em vista que os animais não apresentam a via shikimate. Para vender ainda mais do seu herbicida, a Monsanto encarregou seus geneticistas de plantas de modificarem diversas plantas de cultivo, incluindo milho, para se tornarem resistentes ao glifosato. Para tanto, os cientistas precisaram introduzir uma enzima EPSPS que era resistente à inibição pelo glifosato em plantas de cultivo e em seguida testaram as plantas transformadas em relação à resistência ao herbicida.

Imagine que você é um desses cientistas e que você conseguiu introduzir com sucesso o gene EPSPS resistente nos cromossomos do milho. Você observa que algumas das plantas transgênicas são resistentes ao herbicida, enquanto outras não são. Seu supervisor está muito aborrecido e exige uma explicação sobre o motivo de algumas das plantas não serem resistentes, embora apresentem o transgene em seus cromossomos. Desenhe uma ilustração para ajudá-lo a entender.

31. Muitos cânceres humanos resultam de quando um gene normal sofre mutação e leva ao crescimento descontrolado (um tumor). Os genes que causam câncer, quando mutados, são denominados oncogenes. A quimioterapia é eficaz contra muitos tumores, tendo em vista que tem por alvo células em rápida divisão e mata-as. Infelizmente, a quimioterapia apresenta muitos efeitos colaterais, tais como perda de cabelos ou náuseas, tendo em vista que ela também mata muitas de nossas células normais que estão em rápida divisão, tais como aquelas dos folículos pilosos ou do

revestimento estomacal.

Muitos cientistas e grandes empresas farmacêuticas estão animados a respeito das perspectivas de explorar a via do RNAi para inibir seletivamente os oncogenes em tumores de risco à vida. Explique em termos muito gerais como a terapia com silenciamento gênico pode atuar para tratar o câncer e por que esse tipo de terapia apresentaria menos efeitos colaterais do que a quimioterapia.

32. Você espera que íntrons de autorremoção sejam em média mais longos ou mais curtos que os íntrons recompostos por spliceossomos? Justifique sua resposta.

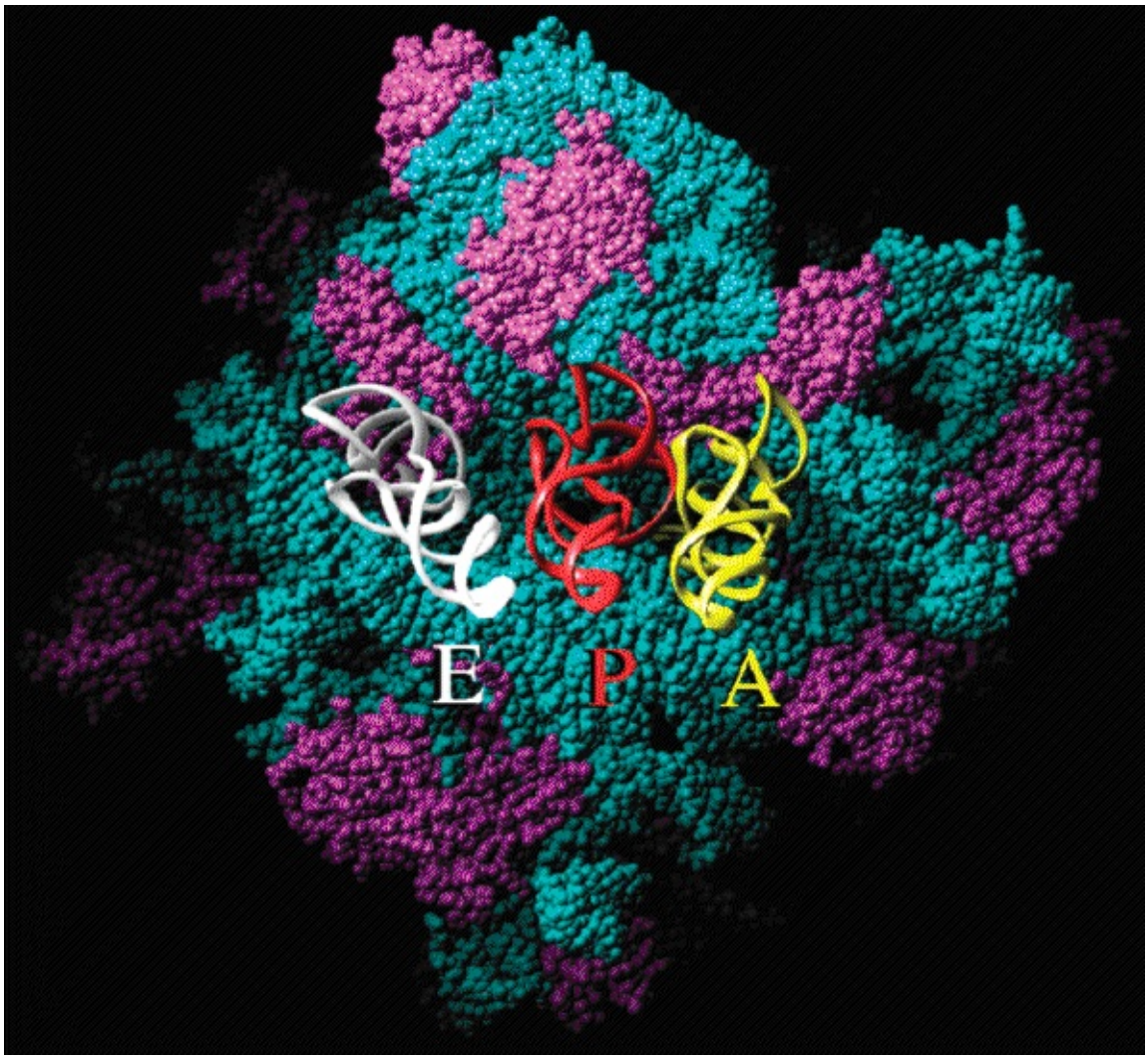
33. Um cientista que inseriu um gene de planta nos cromossomos humanos não foi capaz de detectar qualquer transcrição do gene da planta. Proponha uma explicação com base no que você aprendeu a respeito da transcrição. Agora planeje um experimento para testar sua hipótese.

[1]L. T. Chow, R. E. Gelinas, T. R. Broker e R. J. Roberts, *Cell* 12:1-8, 1977.

CAPÍTULO

9

Proteínas e sua Síntese



Esta imagem demonstra em resolução atômica uma superfície do ribossomo da bactéria *Haloarcula marismortui*, deduzida a partir de cristalografia de raios X. A parte do ribossomo composta por RNA está demonstrada em *azul*; aquela que é composta por proteínas está demonstrada em *roxo*. As estruturas *branca*, *vermelha* e *amarela* no centro são tRNA nos sítios de ligação E, P e A, com suas hastes aceptoras

desaparecendo em uma fissura no ribossomo. (P. Nissen, J. Hansen, N. Ban, P. B. Moore e T. A. Steitz, “The Structural Basis of Ribosome Activity in Peptide Bond Synthesis”, *Science* 289, 2000, 920-930, Fig. 10A. Reimpressa com permissão de AAAS.)

TÓPICOS

9.1 Estrutura proteica
9.2 Código genético
9.3 tRNA | O adaptador
9.4 Ribossomos
9.5 Proteoma

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Comparar as sequências de um gene e sua proteína para avaliar a sua relação
- Examinar e explicar os achados experimentais de que o código genético é não sobreposto e degenerado
- Compreender que embora a tradução seja conservada em todos os seres vivos, existem algumas diferenças importantes entre procariotos e eucariotos
- Comparar e diferenciar os papéis críticos desempenhados por dois RNA funcionais — o RNA ribossômico e o tRNA — na síntese proteica
- Fornecer evidências de que o RNA ribossômico, e não as proteínas ribossômicas, realiza as etapas-chave na tradução
- Comparar diferentes tipos de processamento pós-tradução e sua importância para o funcionamento das proteínas.

Em um pronunciamento no Congresso, em 1969, William Stewart, Ministro da Saúde dos EUA, disse: “É tempo de fechar o livro sobre as doenças infecciosas. A guerra contra a pestilência acabou.” Na ocasião, sua proclamação de vitória não era uma ostentação insensata. Nas duas décadas

anteriores, três doenças infecciosas que haviam atormentado a humanidade durante séculos — poliomielite, varíola e tuberculose — haviam sido notadamente eliminadas em todo o mundo. Um fator importante que contribuiu para a erradicação da tuberculose e de algumas outras doenças infecciosas foi a descoberta e o uso disseminado dos antibióticos, um grupo diverso de compostos químicos que matam patógenos bacterianos específicos sem prejudicar o hospedeiro animal. Antibióticos tais como penicilina, tetraciclina, ampicilina e cloranfenicol, para citar alguns, salvaram centenas de milhões de vidas.

Infelizmente, a proclamação da vitória de William Stewart na batalha contra as doenças infecciosas era prematura. A utilização excessiva de antibióticos mundialmente estimulou a evolução de linhagens bacterianas resistentes. Por exemplo, a cada ano, mais de 2 milhões de pacientes hospitalares nos EUA adquirem uma infecção que é resistente a antibióticos e 90.000 morrem como resultado. Como a resistência se desenvolveu tão rapidamente? As doenças infecciosas serão, novamente, uma causa significativa de mortalidade humana? Ou os cientistas serão capazes de utilizar a sua compreensão sobre os mecanismos de resistência para desenvolver antibióticos mais duradouros?

Para responder a essas questões, cientistas enfocaram no maquinário celular que é o alvo dos antibióticos. Mais da metade de todos os antibióticos atualmente em uso têm por alvo o ribossomo bacteriano, o sítio da síntese proteica em procariotos. Neste capítulo, você aprenderá que os cientistas obtiveram um sucesso incrível com a utilização de uma técnica denominada cristalografia de raios X para visualizar os RNA ribossômicos (rRNA) e as aproximadamente 50 proteínas que compõem as subunidades maior e menor dos ribossomos bacterianos. Embora os ribossomos dos procariotos e eucariotos sejam muito semelhantes, ainda existem diferenças sutis. Em virtude dessas diferenças, os antibióticos são capazes de atacar os ribossomos bacterianos, mas deixam os ribossomos eucarióticos intocados. Com a utilização da cristalografia de raios X, os cientistas também obtiveram sucesso na visualização de antibióticos ligados ao ribossomo (Figura 9.1). A partir desses estudos, eles determinaram que mutações no rRNA bacteriano e/ou em proteínas ribossômicas são responsáveis pela resistência aos antibióticos. Com esse conhecimento sobre os pontos de contato

entre determinados antibióticos e o ribossomo, projetistas de fármacos estão tentando eleborar uma nova geração de antibióticos que, por exemplo, seja capaz de se ligar a múltiplos sítios próximos. Haveria menos probabilidade de desenvolvimento de resistência a um referido fármaco, tendo em vista que seria necessária a ocorrência de duas mutações, o que é um evento muito improvável, mesmo para as bactérias.

Os Capítulos 7 e 8 descreveram como o DNA é copiado de geração para geração e como o RNA é sintetizado a partir de regiões específicas do DNA. Podemos pensar a respeito desses processos como dois estágios da transferência de informação: a *replicação* (a síntese do DNA) e a *transcrição* (a síntese de uma cópia do RNA de uma parte do DNA). Neste capítulo, você aprenderá a respeito do estágio final da transferência de informação: a *tradução* (a síntese de um polipeptídio direcionada pela sequência de RNA).

A ligação de uma molécula de um fármaco ao ribossomo evita a tradução

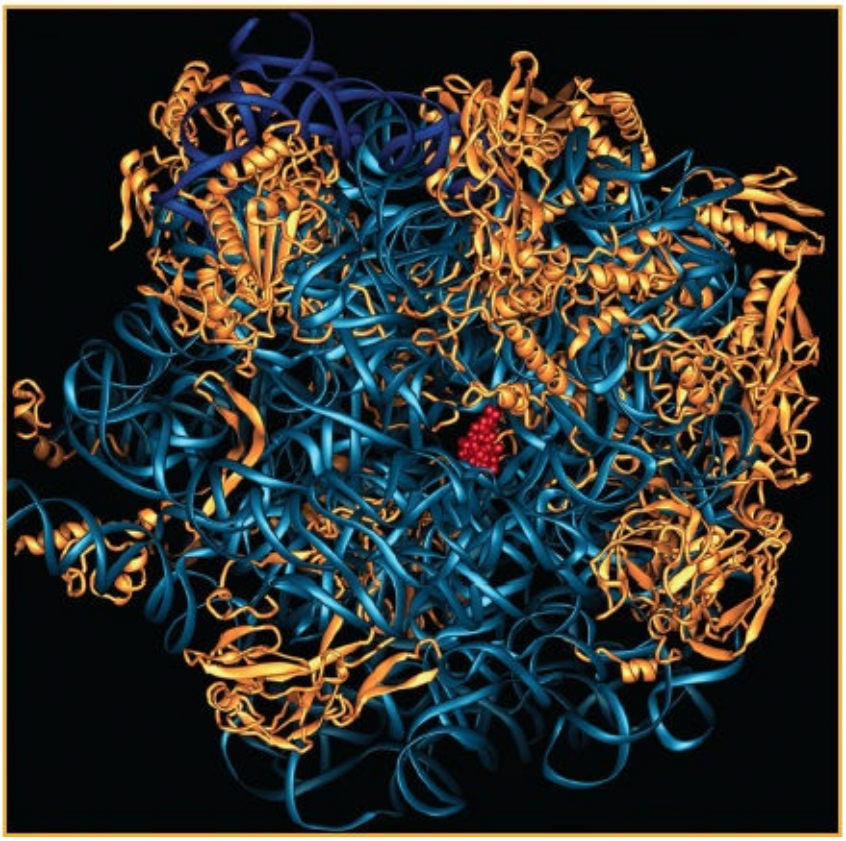


FIGURA 9.1 O fármaco eritromicina (*vermelho*) bloqueia o túnel a partir do qual uma proteína recém-sintetizada surge do ribossomo. A imagem é uma visão superior da subunidade ribossômica 50S na bactéria *Deinococcus radiodurans*. Os RNA ribossômicos estão demonstrados em *azul*, e as proteínas ribossômicas em *amarelo*. (Dr. Joerg Harms, *MPI for Molecular Genetics*, Berlim, Alemanha.)

Como você aprendeu no Capítulo 8, o RNA transcrito a partir dos genes é classificado como RNA mensageiro (mRNA) ou RNA funcional. Neste capítulo, veremos o destino de ambas as classes de RNA. A vasta maioria dos genes codifica mRNA cuja função é atuar como um intermediário na síntese do produto gênico final, a proteína. Contrariamente, relembre que os RNA *funcionais* são ativos como RNA; eles nunca são traduzidos em proteínas. As principais classes de RNA funcionais são atores importantes na síntese proteica. Elas incluem os RNA transportadores (ou de transferência) e os RNA ribossômicos.

- As moléculas de **RNA transportador**, ou de transferência **(tRNA)**, são os adaptadores que traduzem o códon de três nucleotídios no mRNA em um aminoácido correspondente, que é trazido pelo tRNA até o ribossomo no processo de tradução. Os tRNA são componentes gerais do maquinário de tradução; uma molécula de tRNA pode trazer um aminoácido até o ribossomo para fins de traduzir *qualquer* RNA
- Os **RNA ribossômicos (rRNA)** são os principais componentes dos **ribossomos**, que são grandes complexos macromoleculares que juntam aminoácidos para formar a proteína cuja sequência é codificada em um mRNA específico. Os ribossomos são compostos por diversos tipos de rRNA e dezenas de diferentes proteínas. Assim como o tRNA, os ribossomos têm uma função geral, no sentido em que podem ser utilizados para traduzir os mRNA de *qualquer* gene codificador de proteína.

Embora a maioria dos genes codifique mRNA, os RNA funcionais constituem, sem dúvida, a maior fração do RNA celular total. Em uma célula eucariótica típica ativamente em divisão, o rRNA e o tRNA são responsáveis por quase 95% do RNA total, enquanto o mRNA é responsável por apenas aproximadamente 5%. Dois fatores explicam a abundância de rRNA e tRNA. Primeiramente, eles são muito mais estáveis do que os mRNA e, assim, essas moléculas permanecem

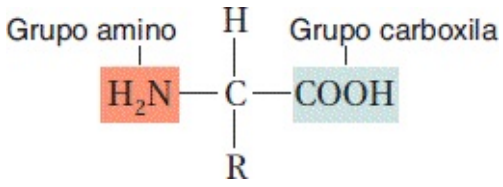
intactas por muito mais tempo. Em segundo lugar, tendo em vista que uma célula eucariótica ativamente em divisão apresenta dezenas de milhares de ribossomos, a transcrição dos genes de rRNA e tRNA constitui mais da metade da transcrição nuclear total nas células eucarióticas ativas e quase 80% da transcrição em células de leveduras.

Os componentes do maquinário de tradução e o processo de tradução são muito semelhantes em procariotos e eucariotos. A principal característica que distingue a tradução em procariotos daquela em eucariotos é o local no qual a transcrição e a tradução ocorrem na célula: os dois processos ocorrem no mesmo compartimento em procariotos, enquanto estão fisicamente separados em eucariotos pela membrana nuclear. Após substancial processamento, os mRNA eucarióticos são exportados a partir do núcleo para a tradução nos ribossomos que estão localizados no citoplasma. Contrariamente, a transcrição e a tradução são acopladas em procariotos: a tradução de um RNA tem início na sua extremidade 5′, enquanto o restante do mRNA ainda está sendo transcrito.

9.1 Estrutura proteica

Quando um transcrito primário é totalmente processado em uma molécula de mRNA madura, ocorre sua tradução em proteína. Antes de considerar como as proteínas são produzidas, precisamos compreender a estrutura proteica. As proteínas são os principais determinantes da forma e da função biológica. Essas moléculas influenciam fortemente a forma, a cor, o tamanho, o comportamento e a fisiologia dos organismos. Tendo em vista que os genes codificam proteínas, compreender a natureza das proteínas é essencial para entender a ação gênica.

Uma proteína é um polímero composto por monômeros denominados **aminoácidos**. Em outras palavras, uma proteína é uma cadeia de aminoácidos. Tendo em vista que os aminoácidos chegaram a ser denominados *peptídios*, a cadeia por vezes é denominada **polipeptídio**. Todos os aminoácidos apresentam a fórmula geral:



Todos os aminoácidos apresentam dois grupos funcionais (carboxila e amino, demonstrados anteriormente) ligados ao mesmo átomo de carbono (denominado carbono α). Também ligados ao carbono α estão um átomo de H e uma cadeia lateral, ou grupo R (reativo). Sabidamente existem 20 aminoácidos nas proteínas, cada um apresentando um grupo R diferente que proporciona ao aminoácido suas propriedades únicas. A cadeia lateral pode ser desde um átomo de hidrogênio (como no aminoácido glicina) até um anel complexo (como no aminoácido triptofano). Nas proteínas, os aminoácidos são unidos por ligações covalentes denominadas ligações peptídicas. Uma ligação peptídica é formada pela ligação da **extremidade amino** (NH_2) de um aminoácido com a **extremidade carboxila** (COOH) de outro aminoácido (Figura 9.2). Uma molécula de água é removida durante a reação. Em virtude do modo pelo qual a ligação peptídica é formada, uma cadeia de polipeptídios sempre apresenta uma extremidade amino (NH_2) e uma extremidade carboxila (COOH), conforme demonstrado na Figura 9.2 A.

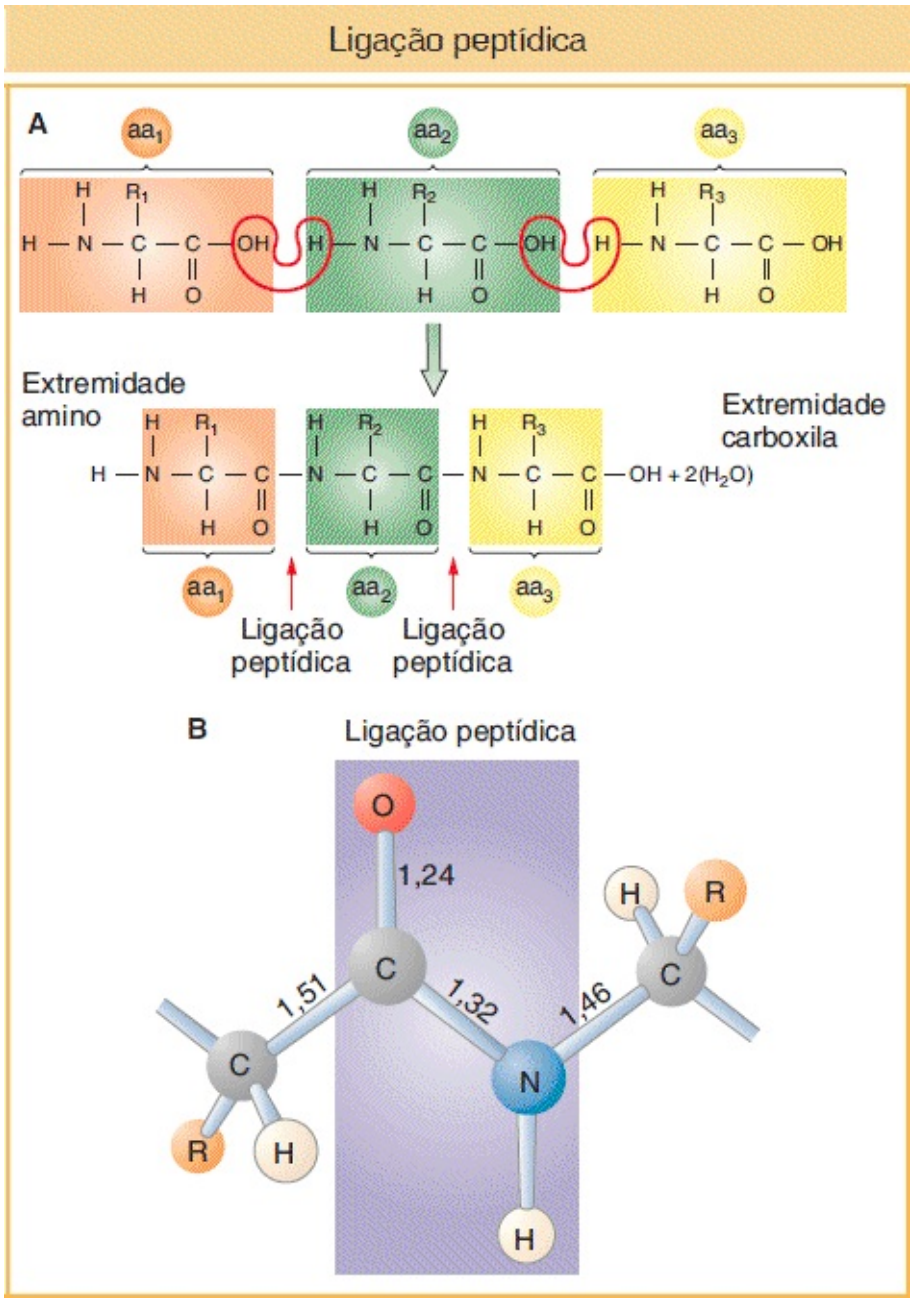


FIGURA 9.2 A. Um polipeptídio é formado pela remoção de água entre os aminoácidos para formar ligações peptídicas. Cada aa indica um aminoácido. R_1, R_2 e R_3 representam os grupos R (cadeias laterais) que diferenciam os aminoácidos. **B.** A ligação peptídica é uma unidade planar rígida, com os grupos R se projetando para fora da espinha dorsal C-N. Estão demonstradas as distâncias das ligações-padrão (em angstroms).

As proteínas apresentam uma estrutura complexa com quatro níveis de organização, ilustrados na Figura 9.3. A sequência linear dos aminoácidos em

uma cadeia de polipeptídios constitui a **estrutura primária** da proteína. Regiões locais da cadeia de polipeptídios se dobram em formas específicas, denominadas **estrutura secundária** da proteína. Cada forma tem origem a partir de forças de ligação entre os aminoácidos que estão próximos na sequência linear. Essas forças incluem diversos tipos de ligações fracas, notavelmente ligações de hidrogênio, forças eletrostáticas e forças de Van Der Waals. As estruturas secundárias mais comuns são a α-hélice e a folha β-pregueada. Diferentes proteínas demonstram uma ou outra, ou por vezes ambas em suas estruturas. A **estrutura terciária** é produzida pelo dobramento da estrutura secundária. Algumas proteínas apresentam **estrutura quaternária**: uma referida proteína é composta por dois ou mais polipeptídios dobrados em separado, também denominados **subunidades**, unidos por ligações fracas. A associação quaternária pode ser entre diferentes tipos de polipeptídios (resultando em um heterodímero se houver duas subunidades), ou entre polipeptídios idênticos (produzindo um homodímero). A hemoglobina é um exemplo de um heterotetrâmero, uma proteína com quatro subunidades; ela é composta por duas cópias de cada um de dois polipeptídios diferentes, demonstrados em verde e em roxo na Figura 9.3 D.

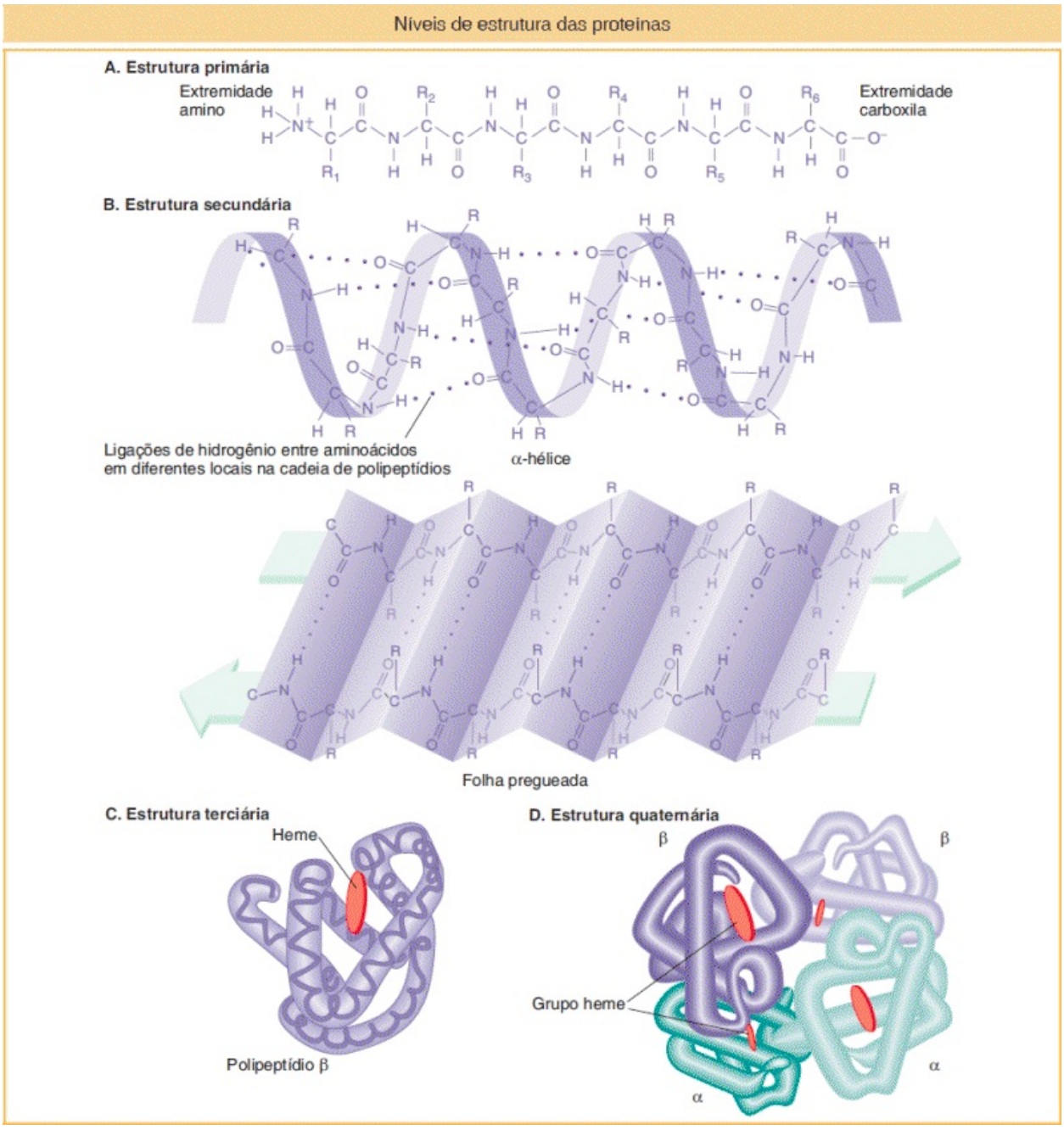


FIGURA 9.3 Uma proteína pode apresentar quatro níveis de estrutura. **A.** Estrutura primária. A sequência de aminoácidos definida por seus grupos R. **B.** Estrutura secundária. O polipeptídio pode formar uma estrutura helicoidal (uma α-hélice) ou uma estrutura em zigue-zague (uma folha β-pregueada). A folha β-pregueada apresenta dois segmentos de polipeptídios dispostos em polaridade oposta, conforme indicado pelas *setas*. **C.** Estrutura terciária. O grupo heme é uma estrutura em anel não proteica, com um átomo de ferro em seu centro. **D.** Estrutura quaternária ilustrada pela hemoglobina, que é composta por quatro subunidades de polipeptídios: duas subunidades α e duas subunidades β.

Muitas proteínas são estruturas compactas; elas são denominadas **proteínas globulares**. As enzimas e os anticorpos estão entre as proteínas globulares mais bem-conhecidas. As proteínas com forma linear, denominadas **proteínas fibrosas**, são componentes importantes de estruturas tais como a pele, os cabelos e os tendões.

A forma é totalmente importante para uma proteína, tendo em vista que a forma específica de uma proteína possibilita que ela realize uma função específica na célula. A forma de uma proteína é determinada pela sua sequência primária de aminoácidos e por condições na célula que promovem o dobramento e a ligação necessários para formar estruturas de nível superior. O dobramento das proteínas em sua conformação correta será discutido no final deste capítulo. A sequência de aminoácidos também determina quais grupos R estão presentes em posições específicas e, portanto, disponíveis para a ligação com outros componentes celulares. Os locais ativos das enzimas são boas ilustrações das interações precisas dos grupos R. Cada enzima apresenta um bolsão denominado **sítio ativo**, ao qual seu substrato ou substratos conseguem se adaptar. No sítio ativo, os grupos R de determinados aminoácidos estão estrategicamente posicionados para interagir com um substrato e catalisar uma reação química específica.

Atualmente, as regras por meio das quais a estrutura primária é convertida em uma estrutura de nível superior não são perfeitamente compreendidas. Entretanto, a partir do conhecimento da sequência primária de aminoácidos de uma proteína, as funções de regiões específicas podem ser previstas. Por exemplo, algumas sequências proteicas características são os pontos de contato com fosfolipídios de membrana que posicionam uma proteína em uma membrana. Outras sequências características atuam para ligar a proteína ao DNA. Sequências de aminoácidos ou dobras nas proteínas que estão associadas a funções em particular são denominadas **domínios**. Uma proteína pode conter um ou mais domínios separados.

9.2 Código genético

A hipótese de um gene–um peptídio de Beadle e Tatum (ver Capítulo 6) foi a fonte da primeira percepção excitante sobre as funções dos genes: os genes de algum modo eram responsáveis pela função das enzimas, e cada gene aparentemente controlava uma enzima. Essa hipótese se tornou um dos grandes conceitos unificadores em biologia, tendo em vista que proporcionou um elemento de ligação dos conceitos e das técnicas de pesquisa de genética e bioquímica. Quando a estrutura do DNA foi deduzida em 1953, aparentava ser provável que houvesse uma correspondência linear entre a sequência de nucleotídios no DNA e a sequência de aminoácidos na proteína. Logo foi deduzido que a sequência de ácido nucleico no mRNA de 5′ para 3′ corresponde à sequência de aminoácidos que vai do N-terminal ao C-terminal.

Se os genes são segmentos de DNA e se um filamento de DNA é apenas uma série de nucleotídios, então a sequência de nucleotídios de algum modo deve ditar a sequência de aminoácidos nas proteínas. Como a sequência de DNA dita a sequência da proteína? A analogia com um código logo vem à mente. A lógica simples nos diz que, se os nucleotídios são as "letras" em um código, então uma combinação de letras pode formar "palavras" que representam diferentes aminoácidos. Primeiramente, é crucial indagar como o código é lido. Ele é sobreposto ou não sobreposto? Em seguida, devemos indagar quantas letras no mRNA formam uma palavra, ou **códon**, e qual ou quais códon(s) representa(m) cada aminoácido. Como decifrar o código genético é a história contada nesta seção.

Códigos sobrepostos versus não sobrepostos

A Figura 9.4 demonstra a diferença entre um código sobreposto e um não sobreposto. O exemplo demonstra um código com três letras, ou **trinca**. Em relação a um código não sobreposto, os aminoácidos consecutivos são especificados por palavras-código consecutivas (códons), conforme demonstrado na parte inferior da Figura 9.4. Em relação a um código sobreposto, os aminoácidos consecutivos são especificados por códons que apresentam algumas bases consecutivas em comum; por exemplo, as duas últimas bases de um códon podem ser também as duas primeiras bases do próximo códon. Códons

sobrepostos estão demonstrados na parte superior da Figura 9.4. Portanto, em relação à sequência AUUGCUCAG em um código não sobreposto, as três trincas AUU, GCU e CAG codificam os três primeiros aminoácidos, respectivamente. Entretanto, em um código sobreposto, as trincas AUU, UUG e UGC codificam os três primeiros aminoácidos se a sobreposição for de duas bases, conforme demonstrado na Figura 9.4.

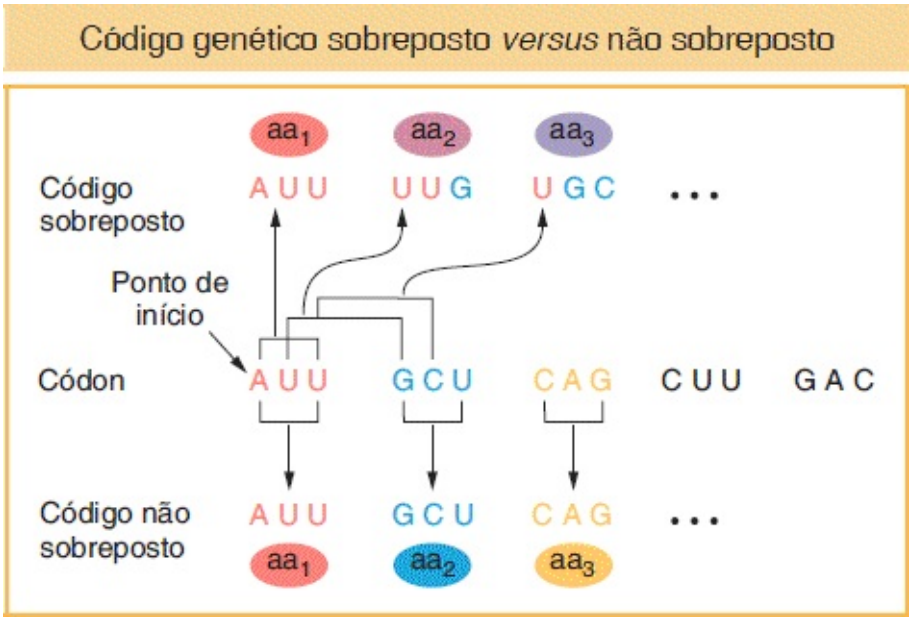


FIGURA 9.4 Um código genético sobreposto e um não sobreposto seriam traduzidos de modo diferente em uma sequência de aminoácidos. O exemplo utiliza um códon com três nucleotídios no RNA (um código triplo). Em um código sobreposto, nucleotídios únicos ocupam posições em múltiplos códons. Nesta ilustração, o terceiro nucleotídio no RNA, U, é encontrado em três códons. Em um código não sobreposto, uma proteína é traduzida por meio da leitura de nucleotídios sequencialmente em conjuntos de três. Um nucleotídio é observado em apenas um códon. Neste exemplo, o terceiro U no RNA está apenas no primeiro códon.

Em 1961, já estava claro que o código genético era não sobreposto. Análises de proteínas alteradas por mutações demonstravam que apenas um único aminoácido é alterado por vez em uma região da proteína. Esse resultado é previsto por um código não sobreposto. Conforme podemos observar na Figura 9.4, um código sobreposto prevê que a alteração de uma única base alterará até três aminoácidos em posições adjacentes na proteína.

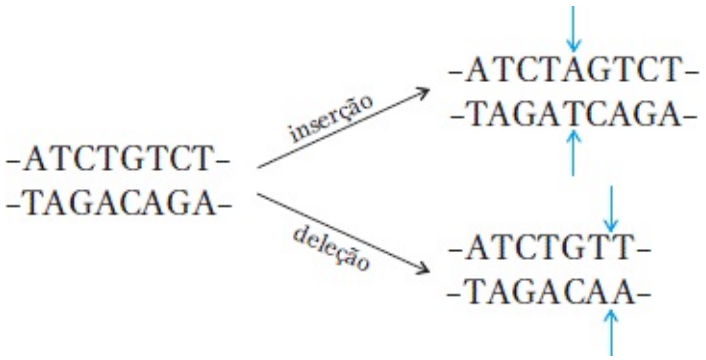
Número de letras no códon

Se uma molécula de mRNA for lida de uma extremidade para a outra, apenas uma das quatro bases diferentes, A, U, G ou C, pode ser observada em cada posição. Portanto, se as palavras que codificam os aminoácidos tivessem uma letra, seriam possíveis apenas quatro palavras. Esse vocabulário não pode ser o código genético, tendo em vista que precisamos ter uma palavra para cada um dos 20 aminoácidos comumente observados nas proteínas celulares. Se as palavras tivessem duas letras, então seriam possíveis $4 \times 4 = 16$ palavras; por exemplo, AU, CU ou CC. Esse vocabulário ainda não é grande o suficiente.

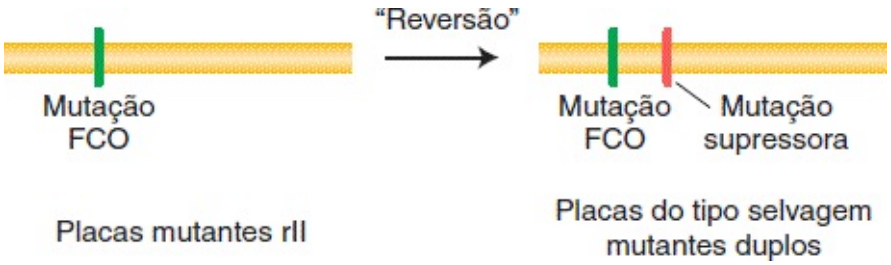
Se as palavras tiverem três letras, então são possíveis $4 \times 4 \times 4 = 64$ palavras; por exemplo, AUU, GCG ou UGC. Esse vocabulário proporciona palavras mais do que suficientes para descrever os aminoácidos. Podemos concluir que as palavras do código devem ser compostas por no mínimo três nucleotídios. Entretanto, se todas as palavras são "trincas", então as palavras possíveis excedem consideravelmente as 20 necessárias para denominar os aminoácidos comuns. Retornaremos a esse excesso de códons posteriormente no capítulo.

Utilização de supressores para demonstrar um código triplo

A comprovação convincente de que um códon tem, de fato, três letras (e não mais do que três) surgiu a partir dos belos experimentos genéticos relatados pela primeira vez em 1961 por Francis Crick, Sidney Brenner e seus colaboradores. Esses experimentos utilizaram mutantes no *locus rII* do fago T4. A utilização de mutações *rII* na análise de recombinação foi discutida no Capítulo 5. O fago T4 normalmente consegue crescer em duas linhagens de *E. coli* diferentes, denominadas B e K. Entretanto, mutações no gene *rII* alteram a gama de hospedeiros do fago: os fagos mutantes ainda podem crescer em um hospedeiro *E. coli* B, mas não conseguem crescer em um hospedeiro *E. coli* K. Mutações que causam esse fenótipo rII foram induzidas por meio da utilização de uma substância química denominada proflavina, que se acreditava atuar por meio da adição ou deleção de pares de nucleotídios únicos no DNA. (Essa presunção tem por base evidências experimentais não apresentadas aqui.) Os exemplos a seguir ilustram a ação da proflavina no DNA bifilamentar.



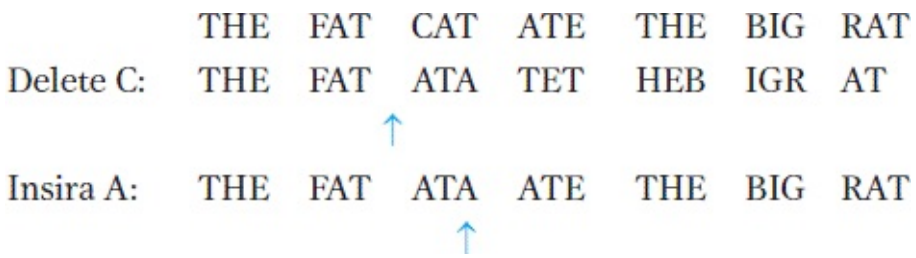
Iniciando com uma mutação específica induzida por proflavina denominada FCO, Crick e seus colegas observaram “reversões” (reversões da mutação) que eram capazes de crescer na linhagem K da *E. coli*. A análise genética dessas placas revelou que os “revertentes” não eram idênticos às de tipos selvagens verdadeiros. Na verdade, observou-se que a reversão ocorre em virtude da existência de uma *segunda mutação* em um sítio diferente daquele de FCO, embora no mesmo gene.



Essa segunda mutação “suprimiu” a expressão mutante de FCO original. Relembre do Capítulo 6 que uma mutação supressora se contrapõe ou suprime os efeitos de outra mutação, de modo que a bactéria é mais semelhante ao tipo selvagem.

Como podemos explicar esses resultados? Se presumirmos que o gene é lido apenas a partir de uma extremidade, então a adição ou deleção original induzida pela proflavina pode resultar em uma mutação, tendo em vista que ela interrompe um mecanismo de leitura normal que estabelece o grupo de bases a ser lido como palavras. Por exemplo, se cada grupo de três bases no mRNA resultante formar uma palavra, então o “quadro de leitura” pode ser estabelecido ao considerar as três primeiras bases a partir da extremidade como a primeira palavra, e as

próximas três como a segunda palavra, e assim por diante. Naquele caso, uma adição ou uma deleção induzida por proflavina de um único par no DNA alteraria o quadro de leitura no mRNA a partir daquele ponto correspondente em diante, causando a leitura errônea de todas as palavras a seguir. Uma referida mutação por deslocamento do quadro de leitura poderia reduzir a maioria da mensagem genética a algo confuso. Entretanto, o quadro de leitura adequado poderia ser restaurado por meio de uma inserção ou uma deleção compensatória em algum outro local, deixando apenas um trecho curto de confusão entre as duas. Considere o exemplo a seguir, no qual palavras de três letras em inglês são utilizadas para representar os códons:



A inserção suprime o efeito da deleção ao restaurar a maior parte do sentido da sentença. Entretanto, por si própria, a inserção também rompe a sentença:

THE FAT CAT AAT ETH EBI GRA T

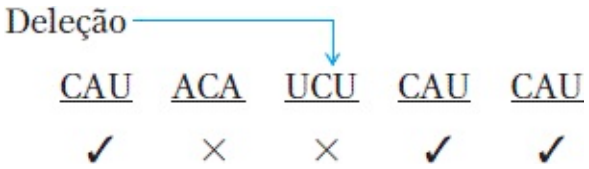
Se presumirmos que o mutante FCO é causado por uma adição, então a segunda mutação (supressora) precisaria ser uma deleção, pois, conforme vimos, apenas uma deleção restauraria o quadro de leitura da mensagem resultante (uma segunda inserção não corrigiria o quadro). Nos diagramas a seguir, utilizamos uma cadeia de nucleotídios hipotética para representar o RNA de maneira simplificada. Também presumimos que as palavras do código apresentam três letras e que são lidas em um sentido (da esquerda para a direita nos nossos diagramas).

1. Mensagem do tipo selvagem.

CAU CAU CAU CAU CAU

2. Mensagem de rII_a: as palavras após a adição são alteradas (×) por deslocamento do quadro de leitura (as palavras marcadas com ✓ não são afetadas).

3. Mensagem de rII_a rII_b: poucas palavras estão erradas, mas o quadro de leitura é restaurado em relação às palavras posteriores.



As poucas palavras erradas no genótipo suprimido podem explicar o fato de que os "revertentes" (fenótipos suprimidos) que Crick e seus colaboradores recuperaram não aparentavam ser de fenótipo exatamente semelhante aos verdadeiros tipos selvagens.

Aqui presumimos que a mudança do quadro de leitura original era uma adição, mas a explicação também funciona bem se presumirmos que a mutação FCO original é uma deleção e o supressor é uma adição. Você pode desejar verificar isso por conta própria. Muito curiosamente, combinações de *três* adições ou *três* deleções demonstraram atuar em conjunto para restaurar um fenótipo do tipo selvagem. Essa observação proporcionou a primeira confirmação experimental de que uma palavra no código genético é composta por três nucleotídios sucessivos, ou uma trinca. O motivo é que três adições ou três deleções em um gene automaticamente restauram o quadro de leitura no mRNA se as palavras forem trincas.

Degeneração do código genético

Conforme já declarado, com quatro letras a partir das quais se pode optar em cada posição, um códon de três letras poderia produzir 4 × 4 × 4 = 64 palavras.

Com apenas 20 palavras necessárias para os 20 aminoácidos comuns, para que são utilizadas as outras palavras, se utilizadas para algo? O trabalho de Crick sugeriu que o código genético é **degenerado**, o que significa que cada uma das 64 trincas deve apresentar algum significado no código. Para que o código seja degenerado, alguns dos aminoácidos devem ser especificados por no mínimo dois ou mais trincas diferentes.

O motivo é o seguinte. Se apenas 20 trincas fossem utilizadas, então as outras 44 não teriam significado, no sentido de que não codificariam qualquer aminoácido. Naquele caso, poderia ser esperado que a maioria das mutações por deslocamento do quadro de leitura produzisse palavras sem sentido, que presumivelmente interromperiam o processo de formação das proteínas, e a supressão das mutações por deslocamento do quadro de leitura raramente, ou nunca, funcionariam. Entretanto, se todas as trincas especificassem algum aminoácido, então as palavras alteradas simplesmente resultariam na inserção dos aminoácidos incorretos na proteína. Portanto, Crick raciocinou que muitos dos ou todos os aminoácidos precisam apresentar diversos nomes diferentes no código de pares de bases; essa hipótese foi posteriormente confirmada bioquimicamente.

CONCEITO-CHAVE Até agora, a discussão demonstra que:

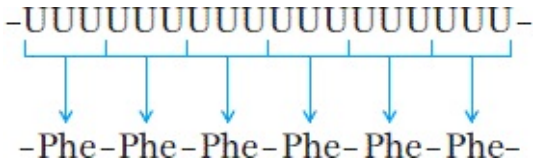
1. A sequência linear de nucleotídios em um gene determina a sequência linear de aminoácidos em uma proteína.
2. O código genético é não sobreposto.
3. Três bases codificam um aminoácido. Essas trincas são denominadas códons.
4. O código é lido a partir de um ponto de início fixo e continua até o término da sequência codificadora. Sabemos que o código é lido sequencialmente, tendo em vista que uma única mutação por deslocamento do quadro de leitura em qualquer parte da sequência codificadora altera o alinhamento do códon em relação ao restante da sequência.
5. O código é degenerado porque alguns aminoácidos são especificados

> por mais de um códon.

Como decifrar o código

Decifrar o código genético — determinar o aminoácido especificado por cada trinca — foi uma das descobertas genéticas mais animadoras dos últimos 50 anos. Após a disponibilização das técnicas experimentais necessárias, o código genético foi rapidamente decifrado.

Um grande avanço foi a descoberta de como produzir mRNA sintético. Se os nucleotídios do RNA são misturados com uma enzima especial (polinucleotídio fosforilase), é formado um RNA unifilamentar na reação. Contrariamente à transcrição, nenhum molde de DNA é necessário para essa síntese e, assim, os nucleotídios são incorporados aleatoriamente. A capacidade de sintetizar RNA ofereceu a excitante perspectiva de criar sequências de mRNA específicas e em seguida observar quais aminoácidos elas especificariam. O primeiro mensageiro sintético obtido foi produzido por meio da mistura apenas de nucleotídios uracila com a enzima de síntese de RNA, produzindo … UUUU … [poli(U)]. Em 1961, Marshall Nirenberg e Heinrich Matthaei misturaram poli(U) com o maquinário de síntese de proteínas de *E. coli in vitro* e *observaram a formação de uma proteína*. A principal empolgação foi centrada na questão da sequência de aminoácidos dessa proteína. Ela comprovou ser a polifenilalanina — uma série de moléculas de fenilalanina ligadas para formar um polipeptídio. Portanto, a trinca UUU deve codificar a fenilalanina:



Por essa descoberta, Nirenberg recebeu o Prêmio Nobel.

Em seguida, foram sintetizados mRNAs que contêm dois tipos de nucleotídios em grupos repetidos. Por exemplo, o mRNA sintético que apresenta a sequência $(AGA)_n$, que é uma longa sequência de AGAAGAAGAAGAAGA, foi utilizado

para estimular a síntese de polipeptídios *in vitro* (em um tubo de ensaio que também continha um extrato celular com todos os componentes necessários para a tradução). A sequência de polipeptídios resultante foi observada a partir de uma diversidade dos referidos testes, com a utilização de diferentes trincas localizadas em outros RNA sintéticos. A partir dos referidos testes, muitas palavras do código puderam ser verificadas. (Esse tipo de experimento está detalhado no Problema 44, ao fim deste capítulo. Ao resolvê-lo, você pode se colocar no lugar de H. Gobind Khorana, que recebeu um Prêmio Nobel por orientar os experimentos.)

Abordagens experimentais adicionais levaram à atribuição de cada aminoácido a um ou mais códons. Relembre que se propôs que o código é degenerado, o que significa que alguns aminoácidos apresentavam mais de um códon determinante. Essa degeneração pode ser observada claramente na Figura 9.5, que fornece os códons e os aminoácidos que eles especificam. Virtualmente todos os organismos sobre a Terra utilizam esse mesmo código genético. (Existem apenas algumas exceções, nas quais um pequeno número de códons apresenta significados diferentes — por exemplo, nos genomas mitocondriais.)

Códons de parada

Você pode ter observado na Figura 9.5 que alguns códons absolutamente não especificam um aminoácido. Esses códons são códons de parada ou término. Eles podem ser considerados como sendo semelhantes aos pontos finais ou às vírgulas que pontuam a mensagem codificada no DNA.

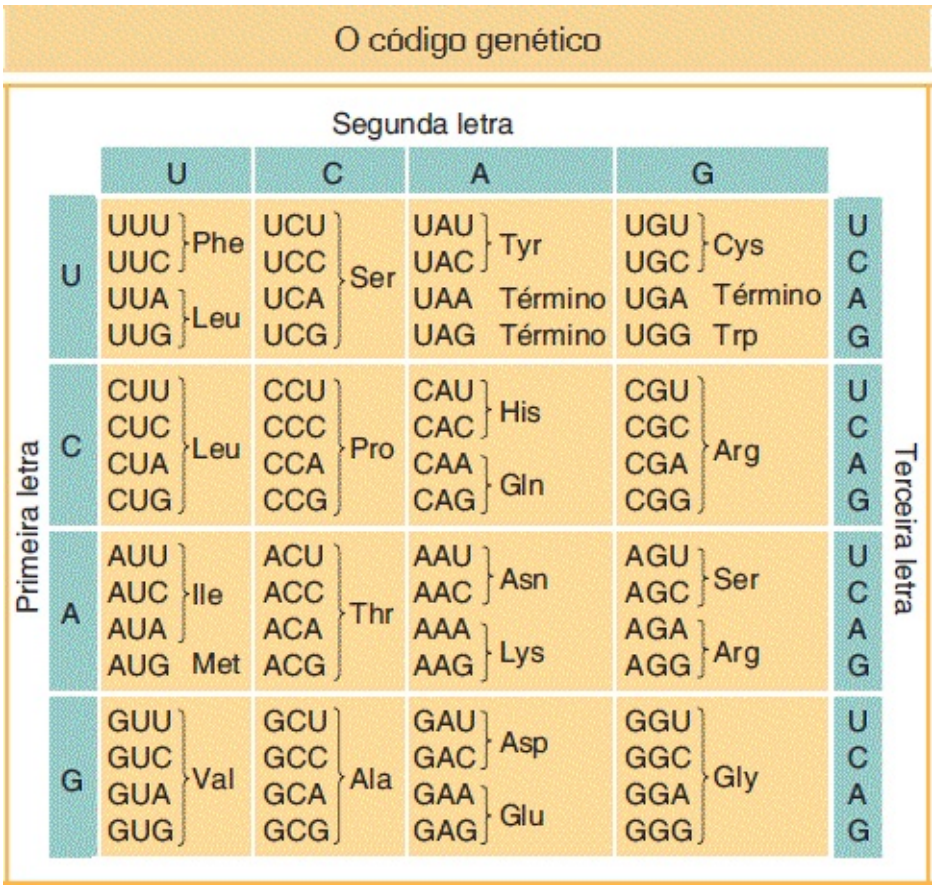


FIGURA 9.5 O código genético designa os aminoácidos especificados por cada códon.

Uma das primeiras indicações da existência de códons de parada surgiu em 1965 a partir do trabalho de Brenner com o fago T4. Brenner analisou determinadas mutações (m_1 a m_6) em um gene único que controla a proteína da cabeça do fago. Ele observou que a proteína da cabeça de cada mutante era uma cadeia polipeptídica mais curta do que aquela do tipo selvagem. Brenner examinou as extremidades das proteínas encurtadas e as comparou à proteína do tipo selvagem. Em relação a cada mutante, ele registrou o próximo aminoácido que *teria* sido inserido para continuar a cadeia do tipo selvagem. Os aminoácidos em relação às seis mutações eram glutamina, lisina, ácido glutâmico, tirosina, triptofano e serina. Esses resultados não apresentam um padrão imediatamente óbvio, mas Brenner deduziu que determinados códons para cada um desses aminoácidos são semelhantes. Especificamente, cada um desses códons pode mutar para o códon UAG por meio de uma única alteração em um par de

nucleotídios do DNA. Portanto, ele postulou que UAG é um códon de parada (término) — um sinal para o mecanismo de tradução de que a proteína agora está completa.

O UAG foi o primeiro códon de parada decifrado; ele é denominado o códon âmbar (âmbar é a tradução do último nome do descobridor do códon, Bernstein). Os mutantes que são defeituosos em virtude da presença de um códon âmbar anormal são denominados *mutantes âmbar*. Dois outros códons de parada são UGA e UAA. De modo análogo ao códon âmbar, e continuando o tema da denominação em relação às cores e às pedras preciosas, UGA é denominado códon opala e UAA é denominado códon ocre. Mutantes que são defeituosos em virtude de conterem códons opala ou ocre anormais são denominados mutantes opala e ocre, respectivamente. Os códons de parada com frequência são denominados códons sem sentido, tendo em vista que não designam aminoácidos.

Além de uma proteína de cabeça mais curta, os fagos mutantes de Brenner apresentavam outra característica interessante em comum: uma mutação supressora (su^-) no cromossomo do hospedeiro causaria o desenvolvimento, pelo fago, de uma proteína de cabeça de comprimento de cadeia normal (tipo selvagem), apesar da presença da mutação *m*. Consideraremos os códons de parada e seus supressores adicionalmente após termos lidado com o processo da síntese proteica.

9.3 tRNA | O adaptador

Após o código genético ter sido decifrado, cientistas começaram a ponderar como a sequência de aminoácidos de uma proteína era determinada pelos códons triplos do mRNA. Um modelo inicial, rapidamente dispensado como ingênuo e improvável, propôs que os códons do mRNA poderiam se dobrar e formar 20 cavidades distintas, que ligam diretamente os aminoácidos específicos na ordem correta. Em vez disto, em 1958, Crick reconheceu o que segue:

> É, portanto, uma hipótese natural que o aminoácido seja levado até o molde por uma molécula adaptadora, e que o adaptador seja a parte que de fato se encaixa no RNA. Em sua forma mais simples, [essa

hipótese] necessitaria de vinte adaptadores, um para cada aminoácido.[1]

Ele especulou que o adaptador "poderia conter nucleotídios. Isso possibilitaria a sua junção ao molde de RNA por meio do mesmo 'pareamento' de bases, conforme observado no DNA". Além disso, "uma enzima em separado seria necessária para unir cada adaptador ao seu próprio aminoácido".

Atualmente, sabemos que a "hipótese do adaptador" de Crick está amplamente correta. De fato, os aminoácidos estão ligados a um adaptador (relembre que os adaptadores constituem uma classe especial de RNA estáveis, denominados *RNA transportador*). Cada aminoácido se torna unido a um tRNA específico, que em seguida transporta aquele aminoácido até o ribossomo, o complexo molecular que ligará o aminoácido a um polipeptídio em crescimento.

Tradução do códon pelo tRNA

A estrutura do RNA mantém o segredo da especificidade entre um códon de mRNA e o aminoácido que ele designa. A molécula de tRNA unifilamentar apresenta uma forma de folha de trevo, composto por quatro hastes de dupla-hélice e três alças unifilamentares (Figura 9.6 A). A alça intermediária de cada tRNA é denominada alça do anticódon, tendo em vista que ela transporta uma trinca de nucleotídios denominada **anticódon**. Essa sequência é complementar ao códon para o aminoácido transportado pelo tRNA. O anticódon no tRNA e o códon no mRNA se ligam por pareamento específico de bases entre os RNA. (Novamente, observamos o princípio da complementaridade de ácidos nucleicos em ação, nesse momento na ligação de dois RNA diferentes.) Tendo em vista que os códons no mRNA são lidos na direção 5′→ 3′, os anticódons estão orientados e escritos na direção 3′→ 5′, conforme a Figura 9.6 A demonstra.

Os aminoácidos são unidos aos tRNA por meio de enzimas denominadas **aminoacil-tRNA sintetases**. Existem 20 dessas enzimas na célula, uma para cada um dos 20 aminoácidos. Cada aminoácido apresenta uma sintetase específica, que o liga apenas àqueles tRNA que reconhecem os códons em relação àquele aminoácido em particular. Para catalisar essa reação, as sintetases apresentam dois sítios de ligação, um para o aminoácido e o outro para o seu tRNA cognato

(Figura 9.7). Um aminoácido é ligado à extremidade 3′ livre de seu tRNA, o aminoácido alanina no caso demonstrado nas Figuras 9.6 A e 9.7. Diz-se que o tRNA com um aminoácido ligado está **carregado**.

Um tRNA normalmente existe como uma folha de trevo dobrada em forma de L, conforme demonstrado na Figura 9.6 B, em vez da folha de trevo "achatada" demonstrada na Figura 9.6 A. A estrutura tridimensional do tRNA foi determinada com a utilização de cristalografia de raios X. Nos anos desde que essa técnica foi utilizada para deduzir a estrutura da dupla-hélice do DNA, ela tem sido refinada, de modo que atualmente pode ser utilizada para determinar a estrutura de macromoléculas muito complexas, tais como o ribossomo. Embora os tRNA difiram em sua sequência primária de nucleotídios, todos os tRNA se dobram virtualmente na mesma conformação em forma de L, com exceção de diferenças na alça do anticódon e na extremidade aminoacil. Essa semelhança de estrutura pode ser facilmente observada na Figura 9.8, que demonstra dois tRNA diferentes sobrepostos. A conservação da estrutura nos mostra que a forma é importante para a função do tRNA.

O que aconteceria se o aminoácido errado fosse covalentemente unido a um tRNA? Um experimento convincente respondeu a essa questão. O experimento utilizou cisteinil-tRNA ($tRNA^{Cys}$), o tRNA específico para cisteína. Esse tRNA foi "carregado" com cisteína, o que significa que a cisteína foi ligada ao tRNA. O tRNA carregado foi tratado com hidreto de níquel, que converteu a cisteína (enquanto ainda estava ligada ao $tRNA^{Cys}$) em outro aminoácido, alanina, sem afetar o tRNA:

$$\text{cisteína-tRNA}^{Cys} \xrightarrow{\text{hidreto de níquel}} \text{alanina-tRNA}^{Cys}$$

A proteína sintetizada com essa espécie híbrida apresentava alanina sempre que era esperada a cisteína. O experimento demonstrou que os aminoácidos são "analfabetos"; eles são inseridos na posição adequada em virtude de os tRNA "adaptadores" reconhecerem os códons do mRNA e inserirem adequadamente seus aminoácidos ligados. Portanto, a ligação do aminoácido correto ao seu tRNA cognato é uma etapa crítica para assegurar que uma proteína seja sintetizada

corretamente. Se o aminoácido errado for ligado, não há como prevenir a sua incorporação em uma cadeia proteica em crescimento.

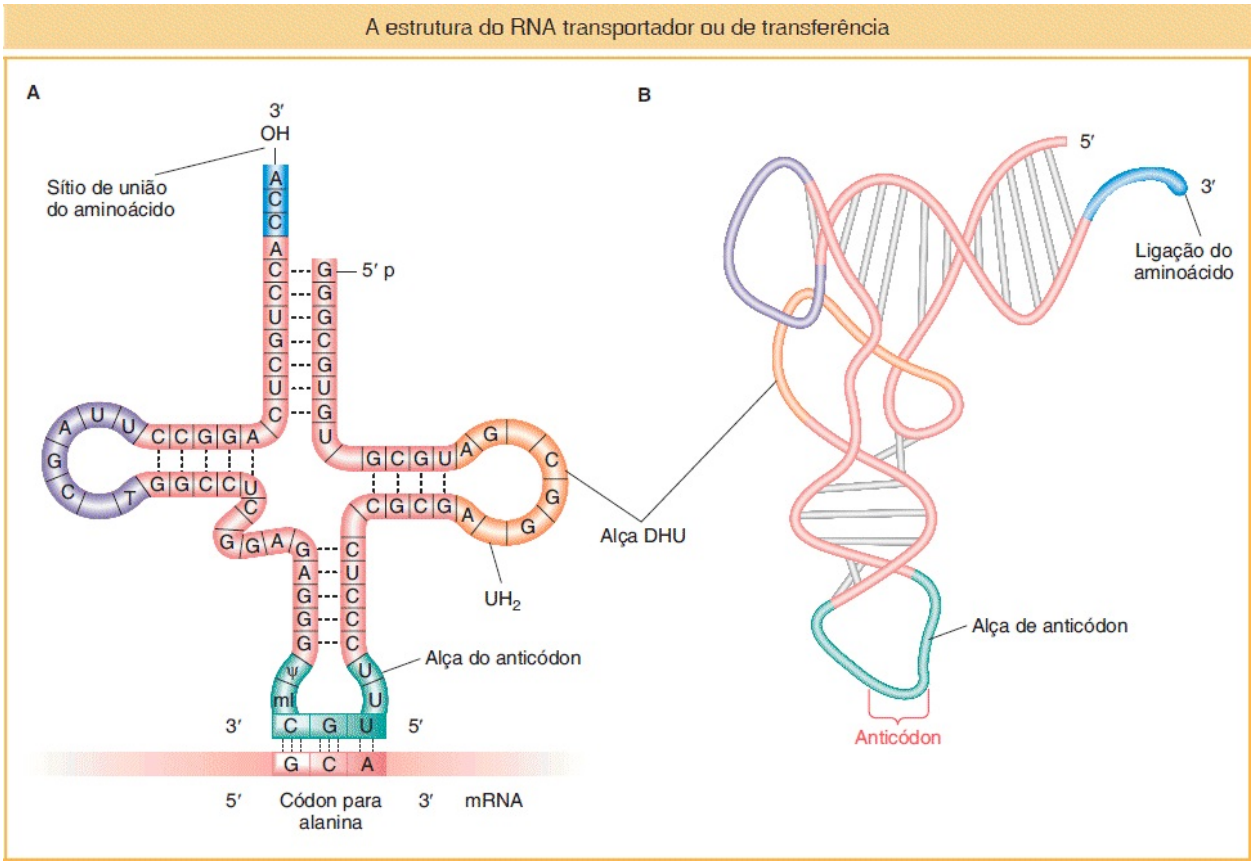


FIGURA 9.6 A. A estrutura do tRNA da alanina de levedura, demonstrando o anticódon do tRNA ligando-se ao seu códon complementar no mRNA. **B.** Diagrama da estrutura tridimensional real do tRNA da fenilalanina de levedura.

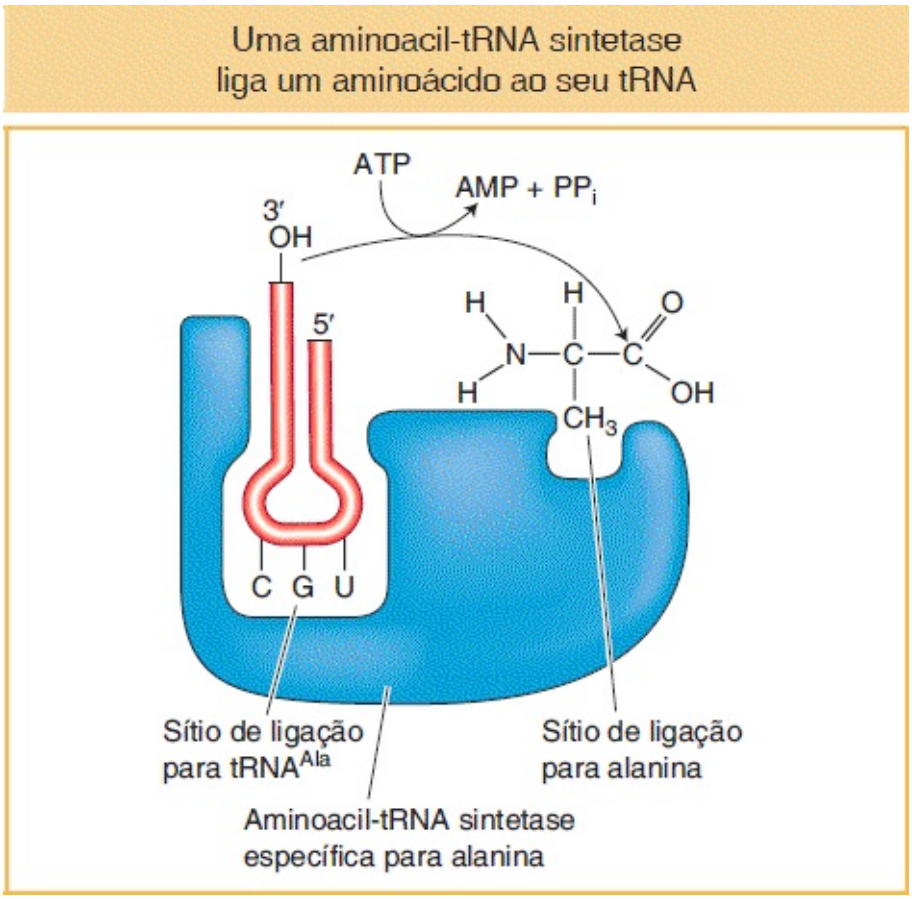


FIGURA 9.7 Cada aminoacil-tRNA sintetase apresenta bolsões de ligação para um aminoácido específico e seu tRNA cognato. Por este meio, um aminoácido é covalentemente ligado ao tRNA com o anticódon correspondente.

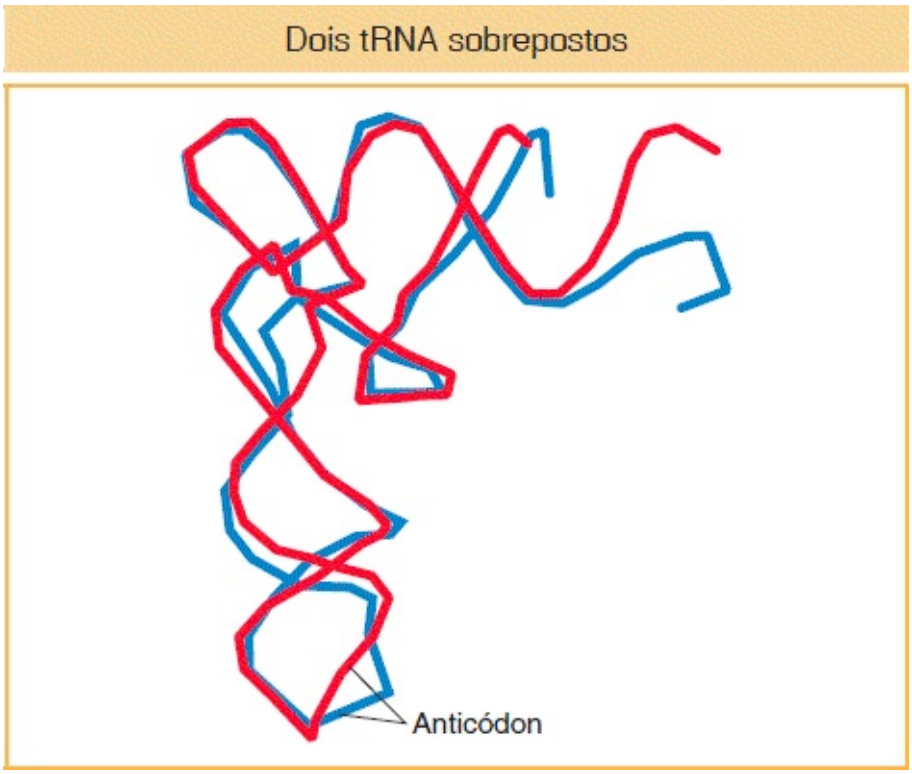


FIGURA 9.8 Quando dobrado em suas estruturas tridimensionais corretas, o tRNA de levedura para glutamina (*azul*) sobrepõe quase completamente o tRNA de levedura para fenilalanina (*vermelho*), exceto em relação à alça do anticódon e à extremidade aminoacil. (Dados de M. A. Rould, J. J. Perona, D. Soll e T. A. Steitz, "Structure of E. coli Glutaminyl–tRNA Synthetase Complexed with tRNA(Gln) and ATP at 2.8 Å Resolution", *Science* 246, 1989, 1135-1142.)

Novamente a degeneração

Conforme pode ser observado na Figura 9.5, o número de códons para um único aminoácido varia desde um códon (UGG em relação ao triptofano) até tantos quantos seis (UCC, UCU, UCA, UCG, AGC ou AGU em relação à serina). O motivo de o código genético conter essa variação não está exatamente claro, mas dois fatores explicam isso:

1. A maioria dos aminoácidos pode ser trazida até o ribossomo por diversos tipos de tRNA alternativos. Cada tipo apresenta um anticódon diferente, que pareia bases com um códon diferente no mRNA.
2. Determinadas espécies de tRNA carregados podem trazer seus aminoácidos específicos para qualquer um de diversos códons. Esses tRNA reconhecem e

se ligam a diversos códons alternativos, não apenas àquele com uma sequência complementar, por meio de um tipo frouxo de pareamento de bases na extremidade 3′ do códon e na extremidade 5′ do anticódon. Esse pareamento frouxo é denominado **oscilante**.

O pareamento oscilante é uma situação na qual o terceiro nucleotídio de um anticódon (na extremidade 5′) pode formar qualquer um de dois alinhamentos (Figura 9.9). Esse terceiro nucleotídio pode formar ligações de hidrogênio com seu nucleotídio complementar normal na terceira posição do códon, ou com um nucleotídio diferente naquela posição. As “regras do pareamento oscilante” ditam quais nucleotídios podem e não podem formar ligações de hidrogênio com nucleotídios alternativos por meio de oscilação (Tabela 9.1). Na Tabela 9.1, a letra *I* faz referência à inosina, uma das raras bases observadas no tRNA, com frequência no anticódon.

Tabela 9.1 Pareamentos de códon-anticódon possibilitados pelas regras de oscilação.

Extremidade 5′ do anticódon	Extremidade 3′ do códon
G	C ou U
C	Apenas G
A	Apenas U
U	A ou G
I	U, C ou A

CONCEITO-CHAVE Diz-se que o código genético é degenerado porque em muitos casos, mais de um códon é atribuído a um único aminoácido; além disso, diversos códons conseguem parear com mais de um anticódon (pareamento oscilante).

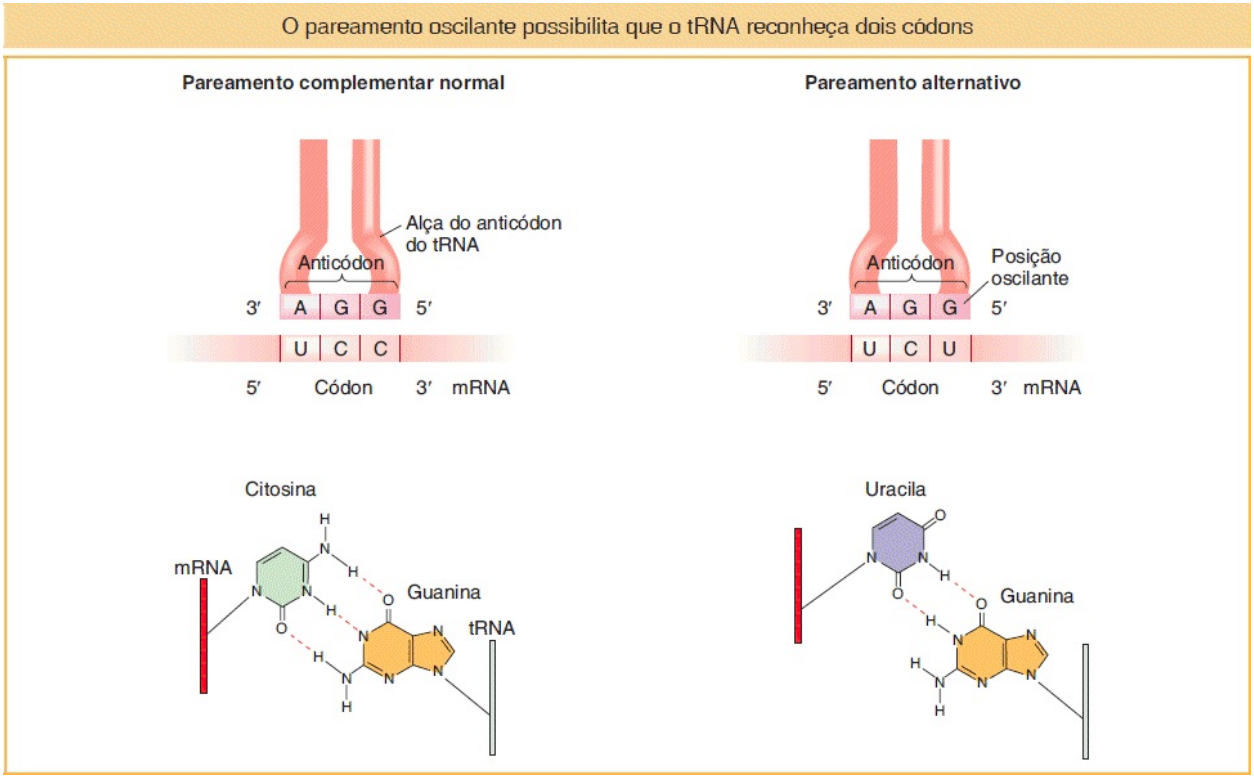


FIGURA 9.9 No terceiro sítio (extremidade 5') do anticódon, G pode adotar qualquer uma de duas posições oscilantes, sendo capaz, portanto, de parear-se com U ou C. Essa capacidade significa que uma única espécie de tRNA que carreia um aminoácido (nesse caso, serina) consegue reconhecer dois códons — UCU e UCC — no mRNA.

9.4 Ribossomos

A síntese proteica ocorre quando moléculas de tRNA e mRNA se associam aos *ribossomos*. A tarefa dos tRNA e do ribossomo é traduzir a sequência de códons de nucleotídios no mRNA na sequência de aminoácidos na proteína. O termo *máquina biológica* foi utilizado nos capítulos antecedentes para caracterizar complexos multissubunitários que realizam funções celulares. O replissomo, por

exemplo, é uma máquina biológica que pode replicar o DNA com precisão e velocidade. O sítio de síntese proteica, o ribossomo, é muito maior e mais complexo do que as máquinas descritas até agora. Sua complexidade ocorre porque ele realiza diversas tarefas com precisão e velocidade. Por isso, é melhor pensar a respeito do ribossomo como uma fábrica que contém muitas máquinas que atuam em conjunto. Vejamos como essa fábrica é organizada para realizar suas diversas funções.

Em todos os organismos, o ribossomo é composto por uma subunidade pequena e uma grande, cada uma composta por RNA (denominado RNA ribossômico ou rRNA) e proteínas. Cada subunidade é composta por um a três tipos de rRNA e até 50 proteínas. As subunidades ribossômicas foram originalmente caracterizadas pela sua velocidade de sedimentação quando centrifugadas em uma ultracentrífuga e, assim, suas denominações são derivadas de seus coeficientes de sedimentação em unidades Svedberg (S), que é uma indicação do tamanho molecular. Em procariotos, as subunidades pequena e grande são denominadas 30S e 50S, respectivamente, e elas se associam para formar uma partícula 70S (Figura 9.10 A). As correspondentes eucarióticas são denominadas 40S e 60S, e o ribossomo completo é denominado 80S (Figura 9.10 B). Embora os ribossomos eucarióticos sejam maiores em virtude de seus componentes maiores e mais numerosos, os componentes e as etapas na síntese proteica são, em geral, semelhantes. As semelhanças indicam claramente que a tradução é um processo antigo, que teve origem em um ancestral comum dos eucariotos e procariotos.

Quando os ribossomos foram estudados pela primeira vez, o fato de que quase dois terços de sua massa são de RNA e apenas um terço é de proteínas foi surpreendente. Durante décadas, havia-se presumido que os rRNA atuavam como andaime, estrutura necessária para a montagem correta das proteínas ribossômicas. Aquele papel aparentava ser lógico, tendo em vista que os rRNA se dobram por meio do pareamento de bases intramolecular em estruturas secundárias estáveis (Figura 9.11). De acordo com esse modelo, as proteínas ribossômicas eram responsáveis apenas pela realização das etapas mais importantes na síntese proteica. A visão foi alterada com a descoberta, na década de 1980, dos RNA catalíticos (ver Capítulo 8). Conforme você verá, atualmente

os cientistas acreditam que os rRNA, assistidos pelas proteínas ribossômicas, realizam a maioria das etapas importantes na síntese proteica.

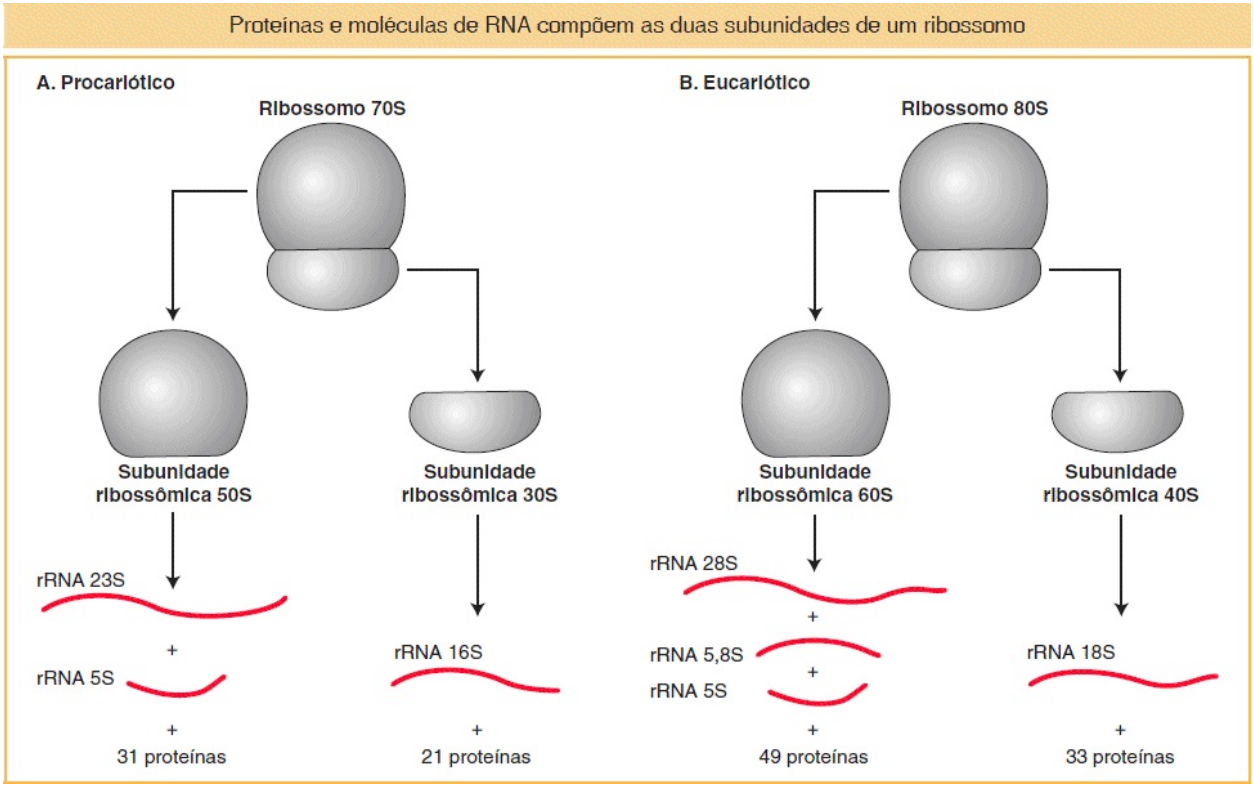


FIGURA 9.10 Um ribossomo contém uma subunidade grande e uma pequena. Cada subunidade contém rRNA de comprimentos variáveis e um conjunto de proteínas. Existem duas moléculas de rRNA principais em todos os ribossomos. Os ribossomos procarióticos também contêm um rRNA de 120 bases que sedimenta em 5S, enquanto os ribossomos eucarióticos apresentam dois rRNA pequenos: uma molécula de RNA 5S semelhante à 5S procariótica e uma molécula 5,8S com 160 bases de comprimento.

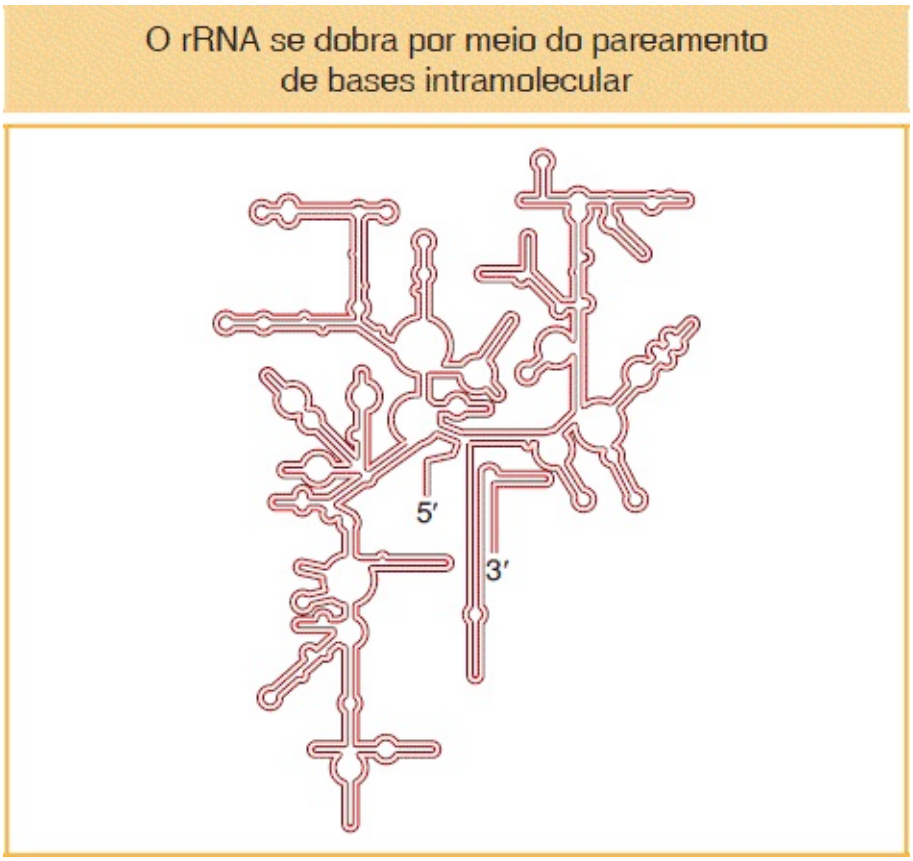


FIGURA 9.11 A estrutura dobrada do RNA ribossômico 16S procariótico da subunidade ribossômica pequena.

Características do ribossomo

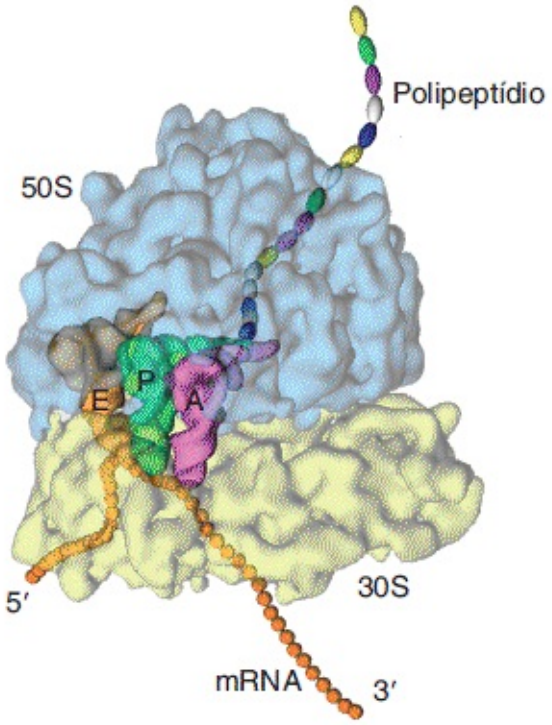
O ribossomo reúne os outros importantes participantes na síntese proteica — as moléculas de tRNA e mRNA — para traduzir a sequência de nucleotídios de um mRNA na sequência de aminoácidos de uma proteína. As moléculas de tRNA e mRNA estão posicionadas no ribossomo de modo que o códon do mRNA consegue interagir com o anticódon do tRNA. Os sítios-chave de interação estão ilustrados na Figura 9.12. O sítio de ligação ao mRNA está completamente dentro da subunidade menor. Existem três sítios de ligação em relação às moléculas de tRNA. Cada tRNA ligado une as subunidades 30S e 50S, posicionado com sua extremidade do anticódon na primeira e sua extremidade aminoacil (que carreia o aminoácido) na última. O **sítio A** (de aminoacil) liga um aminoacil-tRNA que chega cujo anticódon corresponde ao códon no sítio A da subunidade 30S. Na

medida em que prosseguimos na direção de 5′ no mRNA, o próximo códon interage com o anticódon do tRNA no **sítio P** (de peptidil) da subunidade 30S. O tRNA no sítio P liga-se à cadeia de peptídios em crescimento, parte da qual se encaixa em uma estrutura semelhante a um túnel na subunidade 50S. O **sítio E** (de saída) contém um tRNA desacilado (que deixa de carrear um aminoácido) que está pronto para ser liberado do ribossomo. Não está claro se as interações códon-anticódon também ocorrem entre o mRNA e o tRNA no sítio E.

Duas regiões adicionais no ribossomo são críticas para a síntese proteica. O **centro decodificador** na subunidade 30S assegura que apenas os tRNA que carreiam anticódons que correspondem ao códon (denominados tRNA *cognatos*) serão aceitos no sítio A. O **centro peptidiltransferase** na subunidade 50S é o sítio no qual a formação da ligação peptídica é catalisada. Recentemente, muitos laboratórios, especialmente aqueles de Thomas Steitz, Venkatraman Ramakrishnan e Ada Yonath, utilizaram a cristalografia de raios X para "resolver" a estrutura do ribossomo no nível atômico. Por essa conquista, esses três cientistas receberam o Prêmio Nobel de Química em 2009. Os resultados de seus elegantes estudos demonstram claramente que ambos os centros, decodificador e peptidiltransferase, são compostos totalmente por regiões de rRNA; ou seja, os contatos importantes nesses centros são contatos tRNA-rRNA. Acredita-se até que a formação de ligações peptídicas seja catalisada por um sítio ativo no RNA ribossômico e auxiliada apenas por proteínas ribossômicas. Em outras palavras, a subunidade ribossômica maior atua como uma ribozima para catalisar a formação da ligação peptídica.

Sítios-chave de interação no ribossomo

A. Modelo em computador



B. Modelo esquemático

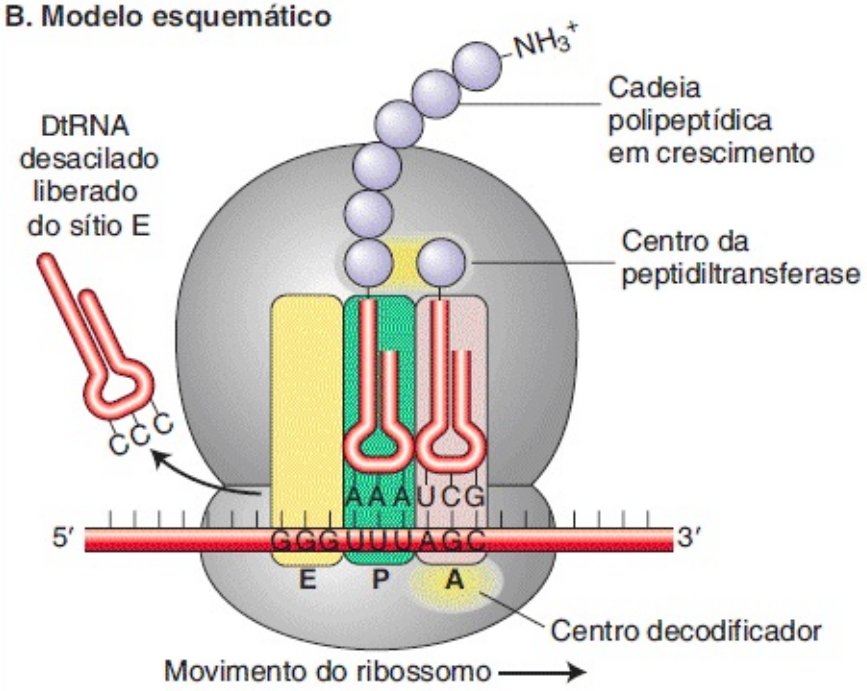


FIGURA 9.12 Sítios-chave de interação em um ribossomo na fase de alongamento da tradução. **A.** Um modelo em computador da estrutura tridimensional do ribossomo, incluindo mRNA, tRNA e cadeia polipeptídica nascente na medida em que ela emerge da subunidade ribossômica maior. **B.** Um modelo esquemático do ribossomo durante o alongamento da tradução. Ver o texto para detalhes. (*A.* J. Frank, Cryo-electron microscopy as an investigative tool: the ribosome as an example, *BioEssays* 23, 2001, 725-732, Figure 2. © Reproduzida, com autorização, de John Wiley & Sons, Inc.)

O QUE OS GENETICISTAS FAZEM ATUALMENTE

Estudos estruturais semelhantes examinaram a subunidade ribossômica grande que forma complexos com diversos antibióticos diferentes. Esses estudos identificaram os pontos de contato entre o antibiótico e o ribossomo e, ao fazer isso, forneceram uma explicação para o motivo de determinados antibióticos inativarem apenas ribossomos bacterianos. Por exemplo, os macrolídeos são uma família de compostos estruturalmente semelhantes que inclui os antibióticos populares eritromicina e Zithromax. Esses antibióticos inibem a síntese proteica parando o ribossomo no mRNA. Eles fazem isso por meio da ligação a uma região específica do rRNA 23S na subunidade ribossômica grande e do bloqueio do assim denominado túnel de saída, onde o polipeptídio nascente emerge dessa subunidade (ver Figura 9.1). Em virtude de pequenas diferenças de sequência entre os rRNA de procariotos e eucariotos, os macrolídios inibem apenas a tradução bacteriana. Curiosamente, as bactérias patogênicas que desenvolveram resistência a alguns desses antibióticos aparentam apresentar mutações ribossômicas que tornam o túnel de saída maior. Portanto, o conhecimento a respeito de como os antibióticos se ligam aos ribossomos auxilia os cientistas a compreenderem como o ribossomo atua e como desenhar novos antibióticos que possam ser ativos contra mutantes resistentes. A utilização de informações básicas a respeito do maquinário celular para o desenvolvimento de novos antibióticos e outros fármacos foi apelidada **desenho de fármaco com base na estrutura**.

Iniciação, alongamento e término da tradução

O processo de tradução pode ser dividido em três fases: iniciação, alongamento e término. Além do ribossomo, do mRNA e dos tRNA, proteínas adicionais são

necessárias para a conclusão bem-sucedida de cada fase. Tendo em vista que determinadas etapas na iniciação diferem significativamente em procariotos e em eucariotos, a iniciação é descrita separadamente em relação aos dois grupos. As fases de alongamento e término são descritas como ocorrem em bactérias, que têm sido o foco de muitos estudos recentes da tradução.

Iniciação da tradução. A principal tarefa da iniciação é posicionar o primeiro aminoacil-tRNA no sítio P do ribossomo e, desse modo, estabelecer o quadro de leitura correto do mRNA. Na maioria dos procariotos e em todos os eucariotos, o primeiro aminoácido em qualquer polipeptídio recém-sintetizado é a metionina, especificada pelo códon AUG. Ela é inserida não pelo $tRNA^{Met}$, mas por um tRNA especial denominado **iniciador**, simbolizado por $tRNA^{Met}_i$. Em bactérias, um grupo formil é adicionado à metionina enquanto o aminoácido se liga ao iniciador, formando a *N*-formilmetionina. (O grupo formil na *N*-formilmetionina é removido posteriormente.)

Como o maquinário de tradução sabe onde iniciar? Em outras palavras, como o códon de iniciação AUG é selecionado entre os muitos códons AUG em uma molécula de mRNA? Relembre que, em ambos os procariotos e os eucariotos, o mRNA apresenta uma região 5′ não traduzida que consiste na sequência entre o sítio de início da transcrição e o sítio de início da tradução (ver Figura 8.7). Conforme você verá posteriormente, a sequência de nucleotídios da 5′ UTR adjacente ao iniciador AUG é crítica para a ligação do ribossomo em procariotos, mas não em eucariotos.

Iniciação em procariotos. Os códons de iniciação são precedidos por sequências especiais denominadas **sequências de Shine-Dalgarno**, que pareiam com a extremidade 3′ de um rRNA, denominado rRNA 16S, na subunidade ribossômica 30S (Figura 9.13). Esse pareamento posiciona corretamente o códon iniciador no sítio P ao qual o tRNA iniciador se ligará. O mRNA pode parear apenas com a subunidade 30S que está dissociada do restante do ribossomo. Observe novamente que o rRNA realiza a função-chave ao assegurar que o ribossomo esteja no local correto para iniciar a tradução.

Três proteínas — IF1, IF2 e IF3 (de **fator de iniciação**) — são necessárias para a iniciação correta (Figura 9.14). A IF3 é necessária para manter a subunidade 30S dissociada da subunidade 50S, e IF1 e IF2 atuam para assegurar que apenas o tRNA iniciador entre no sítio P. A subunidade 30S, o mRNA e o tRNA iniciador constituem o complexo de iniciação. O ribossomo 70S completo é formado por meio da associação da subunidade maior 50S com o complexo de iniciação e da liberação dos fatores de iniciação.

Tendo em vista que um procarioto não apresenta um compartimento nuclear que separe a transcrição e a tradução, o complexo de iniciação procariótico é capaz de se formar em uma sequência de Shine-Dalgarno próxima da extremidade 5′ de um RNA que ainda está sendo transcrito. Portanto, a tradução pode ter início em RNA procarióticos até mesmo antes que eles sejam completamente transcritos.

FIGURA 9.13 Em bactérias, a complementaridade de bases entre a extremidade 3′ do rRNA 16S da subunidade ribossômica pequena (30S) e a sequência de Shine-Dalgarno do mRNA posiciona o ribossomo para iniciar corretamente a tradução no códon AUG downstream.

Iniciação da tradução em procariotos

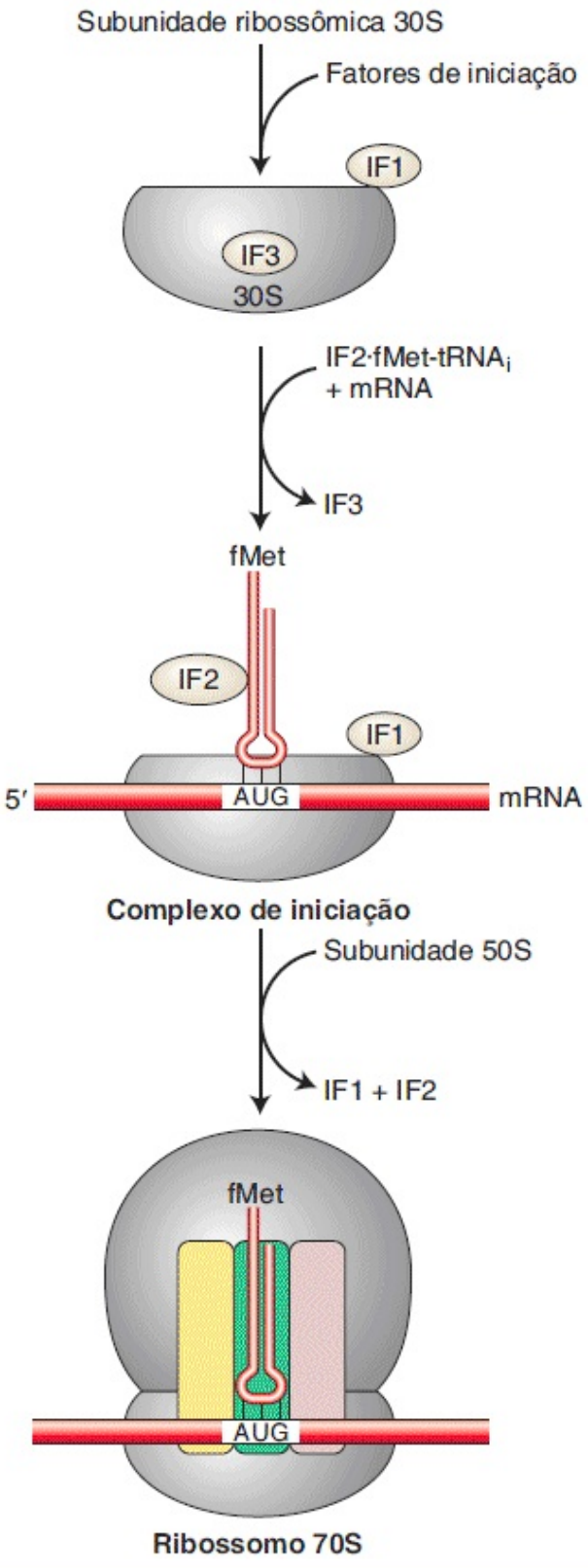


FIGURA 9.14 Os fatores de iniciação auxiliam na montagem do ribossomo no sítio de início da tradução e em seguida se dissociam antes da tradução.

Iniciação em eucariotos. A transcrição e a tradução ocorrem em compartimentos separados na célula eucariótica. Como discutido no Capítulo 8, os mRNA eucarióticos são transcritos e processados no núcleo antes de serem exportados para o citoplasma para a tradução. Na chegada ao citoplasma, o mRNA normalmente é recoberto por proteínas, e algumas regiões podem ter dupla-hélice em virtude do pareamento intramolecular de bases. Essas regiões de estrutura secundária devem ser removidas para expor o códon iniciador AUG. Essa remoção é realizada por fatores de iniciação eucarióticos denominados eIF4A, B e G. Esses fatores de iniciação se associam à estrutura do *cap* (que se encontra na extremidade 5′ de virtualmente todos os mRNA eucarióticos) e à subunidade 40S e ao tRNA iniciador para formar um complexo de iniciação. Uma vez no local, o complexo se movimenta ao longo do mRNA no sentido 5′ para 3′ e desenrola as regiões com bases pareadas (Figura 9.15). Ao mesmo tempo, a sequência exposta é "varrida" à procura de um códon AUG no qual a tradução possa ter início. Após o códon AUG ser alinhado adequadamente com o tRNA iniciador, ao complexo de iniciação se une à subunidade 60S para formar o ribossomo 80S. Assim como em procariotos, os fatores de iniciação eucarióticos se dissociam do ribossomo antes que a fase de alongamento da tradução tenha início.

Iniciação da tradução em eucariotos

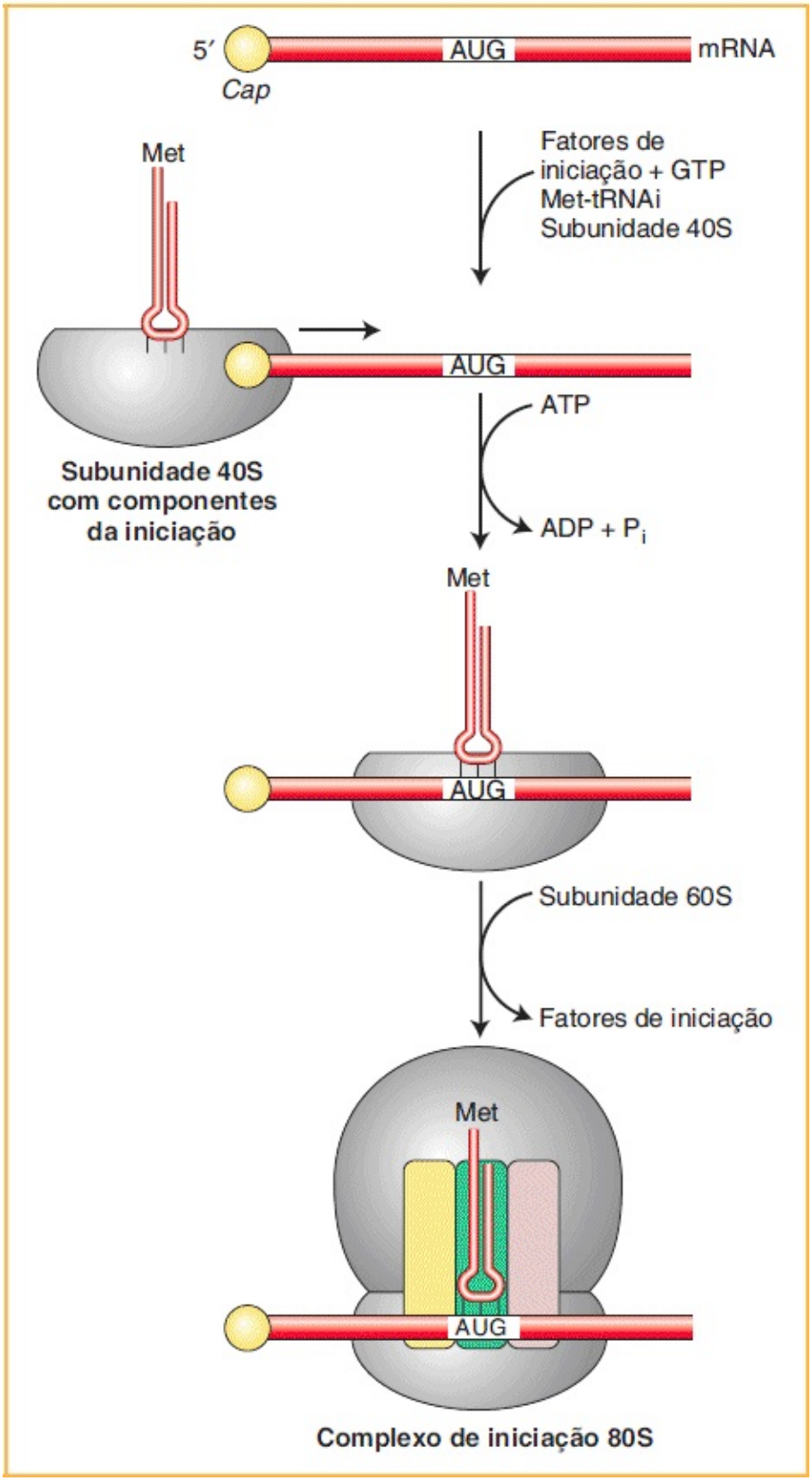


FIGURA 9.15 O complexo de iniciação se forma na extremidade 5′ do mRNA e em seguida realiza a varredura no sentido da extremidade 3′ à procura de um códon de início. O reconhecimento do códon de início aciona a montagem do ribossomo completo e a dissociação dos fatores de iniciação (*não demonstrados*). A hidrólise de ATP fornece energia para direcionar o processo de varredura.

Alongamento. É durante o processo de alongamento que o ribossomo mais se assemelha a uma fábrica. O mRNA atua como um mapa que especifica a entrega de tRNA cognatos, cada um carregando um aminoácido como carga. Cada aminoácido é adicionado à cadeia de polipeptídios em crescimento, enquanto o tRNA desacilado é reciclado por meio da adição de outro aminoácido. A Figura 9.16 detalha as etapas no alongamento. Dois fatores proteicos, denominados fator de alongamento Tu (EF-Tu) e fator de alongamento G (EF-G), auxiliam no processo de alongamento.

Conforme descrito anteriormente neste capítulo, um aminoacil-tRNA é formado por meio da ligação covalente de um aminoácido à extremidade 3′ de um tRNA que contém o anticódon correto. Antes que os aminoacil-tRNA possam ser utilizados na síntese proteica, eles se associam ao fator proteico EF-Tu para formar um complexo ternário composto por tRNA, aminoácido e EF-Tu. O ciclo de alongamento começa com um tRNA iniciador (e sua metionina unida) no sítio P e com o sítio A pronto para aceitar um complexo ternário (ver Figura 9.16). Qual dos 20 complexos ternários diferentes deverá ser aceito é determinado pelo reconhecimento códon-anticódon no centro decodificador da subunidade menor (ver Figura 9.12 B). Quando há correspondência correta, o ribossomo altera sua forma, EF-Tu deixa o complexo ternário e as duas extremidades aminoacil são justapostas no centro peptidiltransferase da subunidade maior (ver Figura 9.12 B). Ali, uma ligação peptídica é formada com a transferência da metionina no sítio P para o aminoácido no sítio A. Nesse ponto, o segundo fator proteico, EF-G, desempenha a sua parte. O fator EF-G aparenta se encaixar no sítio A. A sua entrada nesse sítio altera os tRNA nos sítios A e P para os sítios P e E, respectivamente, e o mRNA se movimenta pelo ribossomo, de modo que o próximo códon é posicionado no sítio A (ver Figura 9.16). Quando o EF-G deixa o ribossomo, o sítio A é aberto para aceitar o próximo complexo ternário.

Nos ciclos subsequentes, o sítio A é preenchido com um novo complexo ternário na medida em que o tRNA desacilado deixa o sítio E. Na medida em que o alongamento progride, o número de aminoácidos no peptidil-tRNA (no sítio P) aumenta. Finalmente, a extremidade amino-terminal do polipeptídio em crescimento emerge do túnel na subunidade 50S e sai do ribossomo.

Término. O ciclo continua até que o códon no sítio A seja um dos três códons de parada: UGA, UAA ou UAG. Relembre que nenhum tRNA reconhece esses códons. Em vez disso, proteínas denominadas **fatores de liberação** (RF1, RF2 e RF3 em bactérias) reconhecem os códons de parada (Figura 9.17). Em bactérias, RF1 reconhece UAA ou UAG, enquanto RF2 reconhece UAA ou UGA; ambos são auxiliados por RF3. A interação dos fatores de liberação 1 e 2 e o sítio A difere daquela do complexo ternário de dois modos importantes. Primeiramente, os códons de parada são reconhecidos por tripeptídios nas proteínas RF, não por um anticódon. Em segundo lugar, os fatores de liberação se encaixam no sítio A da subunidade 30S, mas não participam na formação da ligação peptídica. Em vez disto, uma molécula de água entra no centro peptidiltransferase, e sua presença leva à liberação do polipeptídio do tRNA no sítio P. As subunidades ribossômicas se separam, e a subunidade 30S agora está pronta para formar um novo complexo de iniciação.

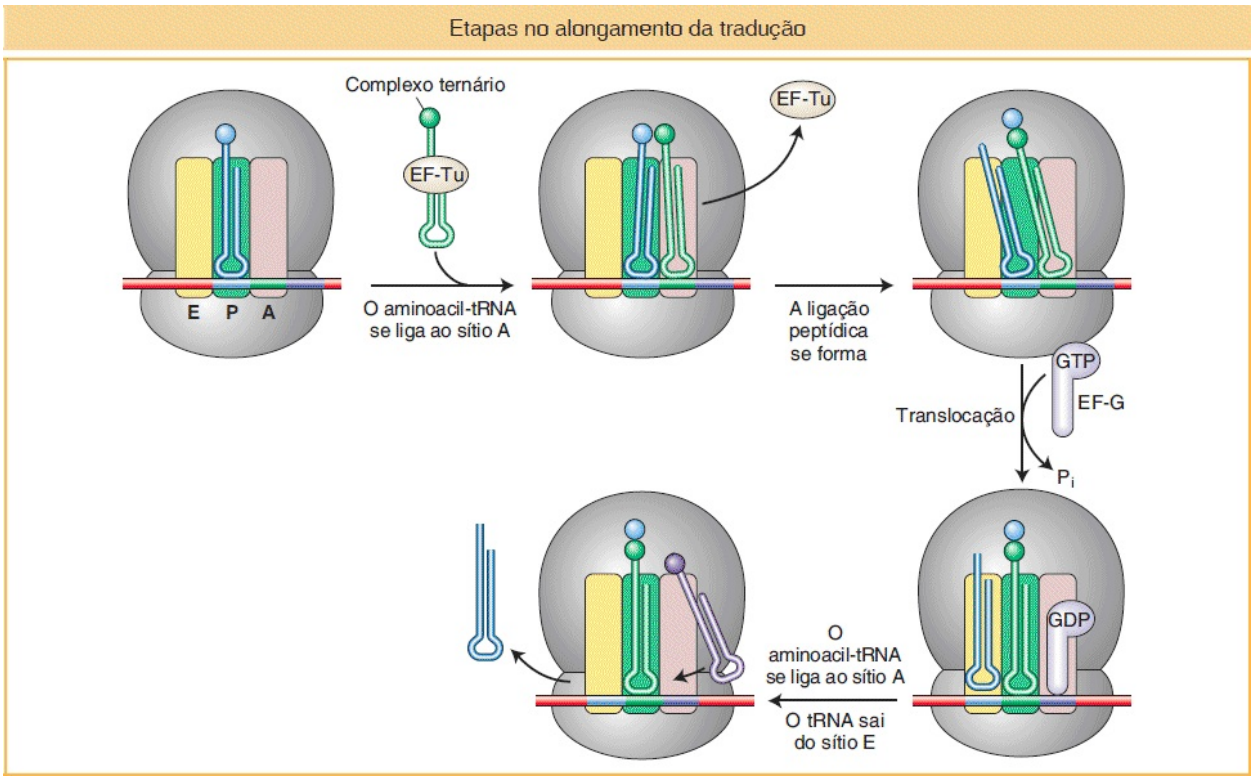


FIGURA 9.16 Um complexo ternário composto por um aminoacil-tRNA ligado a um fator EF-Tu se liga ao sítio A. Quando o seu aminoácido se une à cadeia de polipeptídios em crescimento, um fator EF-G se liga a sítio A enquanto “empurra” os tRNA e seus códons do mRNA para os sítios E e P. Ver o texto para detalhes.

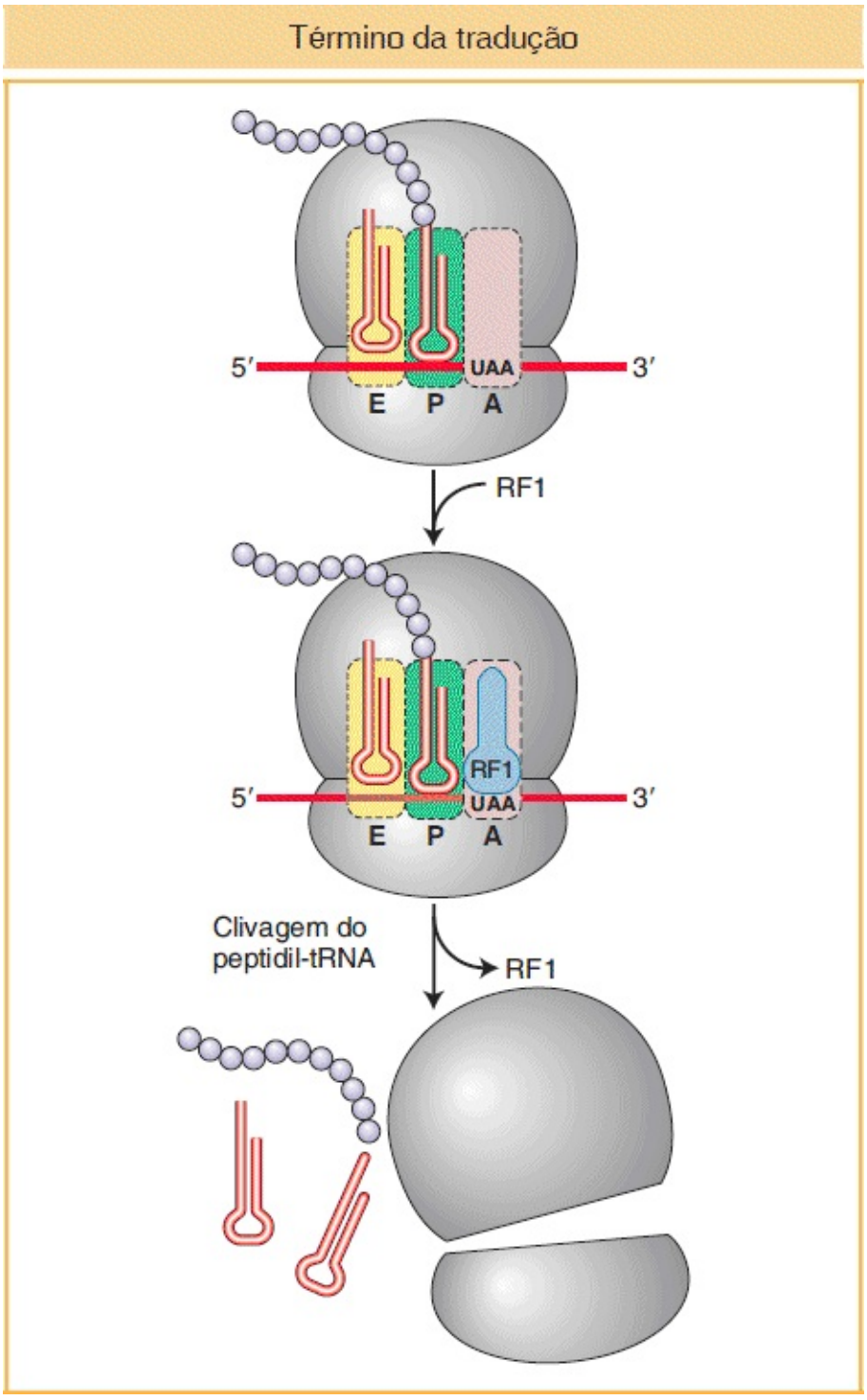


FIGURA 9.17 A tradução é encerrada quando fatores de liberação reconhecem os códons de parada no sítio A do ribossomo.

CONCEITO-CHAVE A tradução é realizada pelos ribossomos que se

> movimentam ao longo do mRNA no sentido 5′ para 3′. Um conjunto de moléculas de tRNA traz os aminoácidos até o ribossomo, e seus anticódons se ligam aos códons do mRNA expostos no ribossomo. Um aminoácido que chega se torna ligado à extremidade amino da cadeia de polipeptídios em crescimento no ribossomo.

Mutações supressoras sem sentido

É interessante considerar os supressores das mutações sem sentido definidas por Brenner e colaboradores. Relembre que as mutações em fagos denominadas mutantes âmbar substituíram os códons do tipo selvagem por códons de parada, mas as mutações supressoras no cromossomo hospedeiro se contrapuseram aos efeitos das mutações âmbar. Agora podemos dizer mais especificamente onde as mutações supressoras estavam localizadas e como elas atuaram.

Muitos desses supressores são mutações em genes que codificam tRNA e são conhecidas como supressoras do tRNA. Essas mutações alteram as alças do anticódon de tRNA específicos de tal modo que um tRNA se torna capaz de reconhecer um códon de parada no mRNA. Na Figura 9.18, a mutação âmbar substitui um códon do tipo selvagem pelo códon de término de cadeia UGA. Por si próprio, o UAG causaria o corte prematuro da proteína na posição correspondente. A mutação supressora nesse caso produz um $tRNA^{Tyr}$ com um anticódon que reconhece o códon de parada UAG mutante. Portanto, no mutante suprimido, o $tRNA^{Tyr}$ compete com o fator de liberação pelo acesso ao códon de parada UAG. Como resultado, se a tirosina for inserida, a tradução continua além daquela trinca.

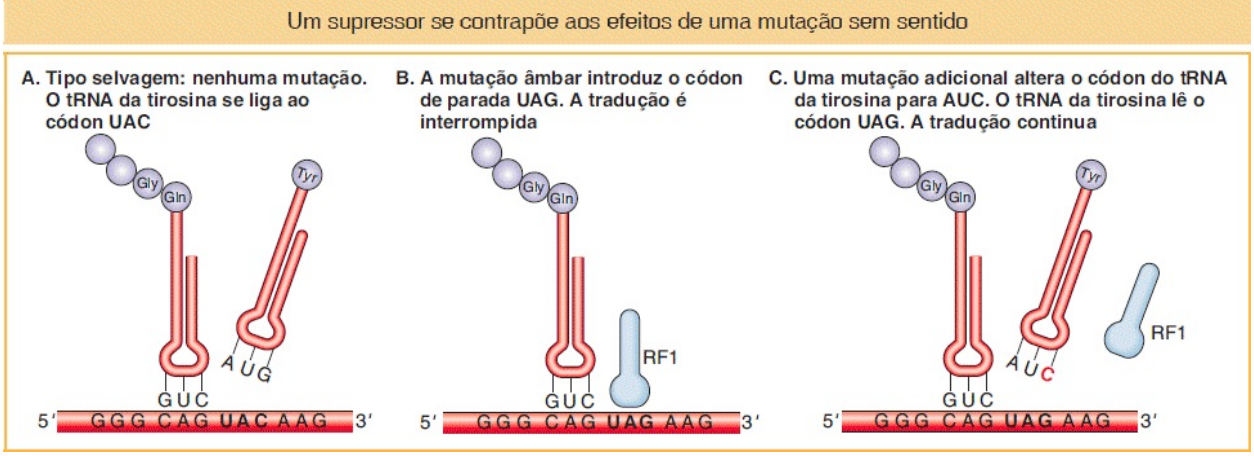


FIGURA 9.18 Um supressor possibilita que a tradução continue enquanto uma mutação a teria interrompido. **A.** No tipo selvagem, um tRNA lê o códon UAC e a tradução continua. **B.** Término da tradução. Aqui, o aparato de tradução não consegue passar por um códon de parada (neste caso, UAG), tendo em vista que nenhum tRNA consegue reconhecer a trinca UAG. Em vez disso, um fator de liberação se liga ao códon e a síntese proteica termina, com a subsequente liberação do fragmento polipeptídico. **C.** Uma mutação altera o anticódon de um tRNA de tirosina, de modo que esse tRNA agora consegue ler o códon UAG. A supressão do códon UAG pelo tRNA alterado agora possibilita o alongamento da cadeia.

Supressoras de tRNA (como $tRNA^{Tyr}$) também poderiam se ligar a sinais de término normais e resultar na síntese de proteínas anormalmente longas? Agora que muitos genomas foram sequenciados, sabe-se que um dos três códons de parada, UAA (ocre), é utilizado com muito mais frequência para terminar a síntese proteica. Como tal, não é surpresa que células com supressores ocre normalmente sejam mais doentes do que células com mutações supressoras âmbar ou opala.

9.5 Proteoma

O Capítulo 8 foi iniciado com uma discussão sobre o número de genes no genoma humano e como aquela número (aproximadamente 21.000) era muito inferior ao número real de proteínas em uma célula humana (mais de 100.000). Agora que você está familiarizado sobre como a informação codificada no DNA é transcrita em RNA e como o RNA é traduzido em proteína, esse é um bom momento para revisar esse assunto e observar de modo mais próximo as fontes de diversificação

das proteínas. Primeiramente, revisaremos alguns poucos termos antigos e adicionaremos um novo, que será útil nessa discussão. Você já sabe que o *genoma* é todo o conjunto de material genético em um organismo. Você aprenderá no Capítulo 14 que o *transcriptoma* é o conjunto completo de transcritos codificadores e não codificadores em um organismo, órgão, tecido ou célula. Outro termo é o **proteoma**, que foi brevemente introduzido no Capítulo 8, mas que é definido aqui como o conjunto completo de proteínas em um organismo, órgão, tecido ou célula. No restante deste capítulo, você verá como o proteoma é enriquecido por dois processos celulares: a recomposição alternativa do pré-mRNA e a modificação pós-tradução das proteínas.

A recomposição alternativa gera isoformas de proteínas

Conforme você relembra do Capítulo 8, a recomposição alternativa do pré-mRNA possibilita que um gene codifique mais de uma proteína. As proteínas são compostas por domínios funcionais que com frequência são codificados por diferentes éxons. Portanto, a recomposição alternativa de um pré-mRNA pode levar à síntese de diversas proteínas (denominadas **isoformas**) com diferentes combinações de domínios funcionais. Esse conceito é ilustrado pelo *FGFR2*, um gene humano que codifica o receptor que liga os fatores de crescimento de fibroblastos e em seguida transduz um sinal dentro da célula (Figura 9.19). A proteína FGFR2 é composta por diversos domínios, incluindo um domínio de ligação a ligante extracelular. A recomposição alternativa resulta em duas isoformas que diferem em seus domínios extracelulares. Em virtude dessa diferença, cada isoforma liga-se a diferentes fatores de crescimento. Para muitos genes que são recompostos alternativamente, diferentes isoformas são produzidas em diferentes tecidos.

Eventos pós-tradução

Quando liberadas do ribossomo, a maioria das proteínas recém-sintetizadas é incapaz de atuar. Conforme você verá nesta seção e nos capítulos subsequentes deste livro, a sequência de DNA é apenas uma parte da história de como os organismos funcionam. Nesse caso, todas as proteínas recém-sintetizadas

precisam se dobrar corretamente e os aminoácidos de algumas proteínas precisam ser quimicamente modificados. Tendo em vista que uma parte do dobramento e da modificação das proteínas ocorre após a síntese proteica, eles são denominados eventos pós-tradução.

Dobramento da proteína dentro da célula. O evento pós-tradução mais importante é o dobramento da proteína nascente (recém-sintetizada) em sua forma tridimensional correta. Diz-se que uma proteína que é dobrada corretamente está em sua conformação nativa (contrariamente a uma proteína não dobrada ou dobrada erroneamente, que é não nativa). Como observamos no início deste capítulo, as proteínas existem em uma extraordinária diversidade de estruturas. As estruturas distintas das proteínas são essenciais para a sua atividade enzimática, para a sua capacidade de se ligar ao DNA, ou para seus papéis estruturais na célula. Embora se saiba desde a década de 1950 que a sequência de aminoácidos de uma proteína determina a sua estrutura tridimensional, também se sabe que o ambiente aquoso dentro da célula não favorece o dobramento correto da maioria das proteínas. Tendo em vista que as proteínas de fato se dobram corretamente na célula, uma questão antiga tem sido *como o dobramento correto é realizado*.

A resposta aparentemente é que as proteínas nascentes são dobradas corretamente com o auxílio de chaperonas — uma classe de proteínas encontrada em todos os organismos, desde bactérias até plantas e seres humanos. Uma família de chaperonas, denominada chaperoninas GroE, forma grandes complexos multissubunitários denominados máquinas de dobramento de chaperoninas. Embora o mecanismo preciso ainda não seja compreendido, acredita-se que proteínas recém-sintetizadas e não dobradas entrem em uma câmara na máquina de dobramento que proporciona um microambiente eletricamente neutro, dentro do qual a proteína nascente consegue se dobrar com sucesso em sua conformação nativa.

Modificação pós-tradução das cadeias laterais de aminoácidos. Conforme já declarado, as proteínas são polímeros de aminoácidos produzidos a partir de

qualquer um dos 20 diferentes tipos. Entretanto, a análise bioquímica de muitas proteínas revela que uma diversidade de moléculas pode ligar-se covalentemente às cadeias laterais de aminoácidos. Mais de 300 modificações de cadeias laterais de aminoácidos são possíveis após a tradução. Duas das modificações pós-tradução mais comumente encontradas — a fosforilação e a ubiquitinação — são consideradas em seguida.

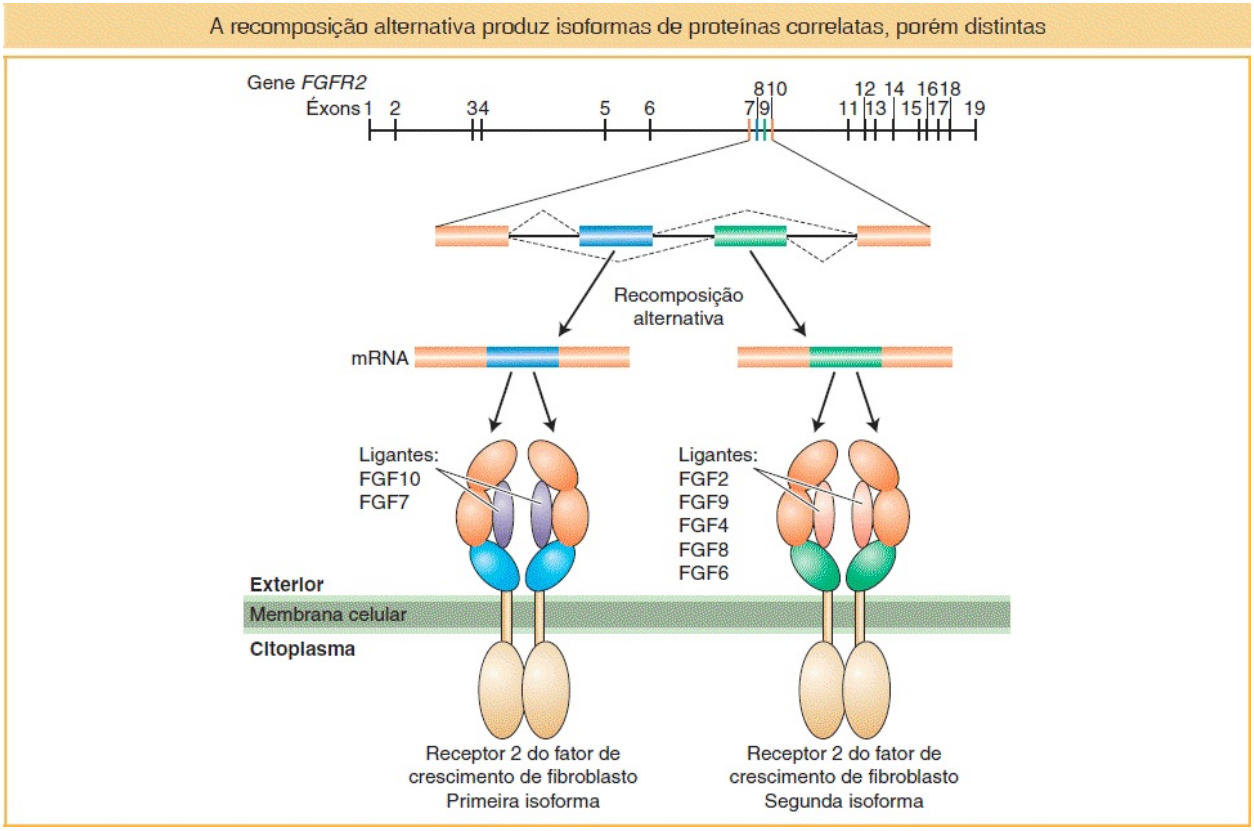


FIGURA 9.19 RNA mensageiros produzidos por recomposição alternativa do pré-mRNA do gene *FGFR2* humano codificam duas isoformas de proteína que se ligam a diferentes ligantes (os fatores de crescimento).

Fosforilação. Enzimas denominadas *quinases* unem os grupos fosfato aos grupos hidroxila dos aminoácidos serina, treonina e tirosina, enquanto enzimas denominadas *fosfatases* removem esses grupos fosfato. Tendo em vista que os grupos fosfato são negativamente carregados, a sua adição a uma proteína normalmente altera a conformação proteica. A adição e a remoção de grupos fosfato atuam como uma alteração reversível para controlar uma diversidade de

eventos celulares, incluindo a atividade enzimática, as interações proteína–proteína e as interações proteína–DNA (Figura 9.20).

Uma medida da importância da fosforilação de proteínas é o número de genes que codificam a atividade quinase no genoma. Até mesmo um organismo simples como uma levedura apresenta centenas de genes de quinases, enquanto a planta da mostarda *Arabidopsis thaliana* apresenta mais de 1.000. Outra medida da significância da fosforilação de proteínas é que a maior parte das diversas interações de proteínas que ocorrem em uma célula típica é regulada por fosforilação.

Análises recentes das interações de proteínas do proteoma indicam que a maioria das proteínas atua por meio da interação com outras proteínas. **Interatoma** é a denominação dada ao conjunto completo de interações de proteínas em um organismo, órgão, tecido, ou célula. Um modo de ilustrar a rede de interações de proteínas que constitui um interatoma está demonstrado na Figura 9.21. Para gerar esta figura, pesquisadores determinaram as 3.186 interações proteicas entre 1.705 proteínas humanas. Entretanto, essas interações constituem apenas uma pequena fração das interações de proteínas que estão ocorrendo em todas as células humanas sob todas as condições de crescimento.

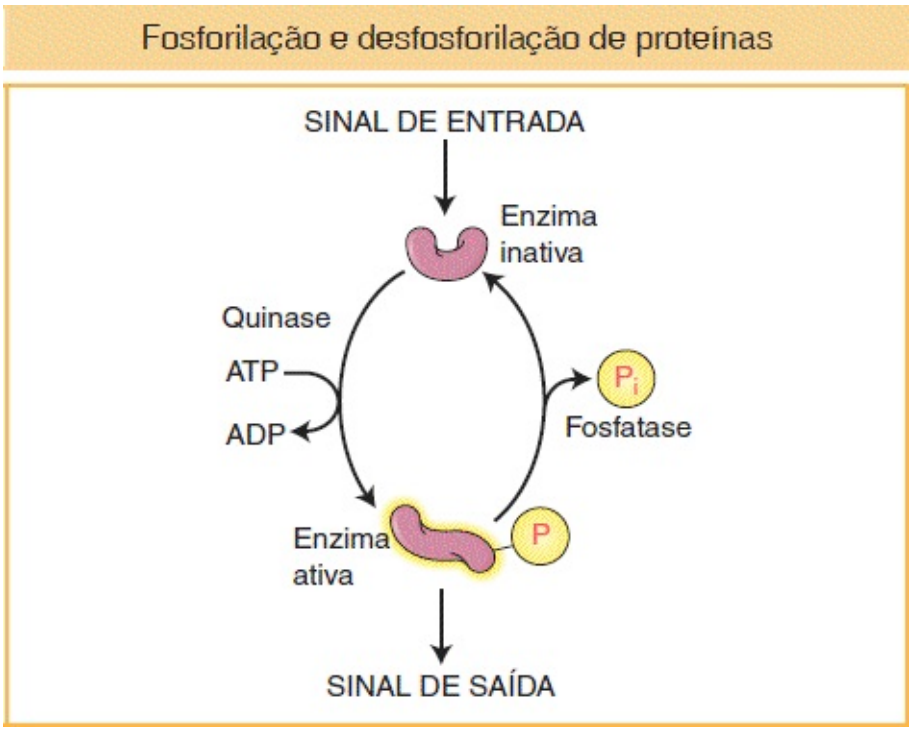


FIGURA 9.20 As proteínas podem ser ativadas por meio da ligação enzimática de grupos fosfato aos grupos laterais de seus aminoácidos e inativadas por meio da remoção desses grupos fosfato.

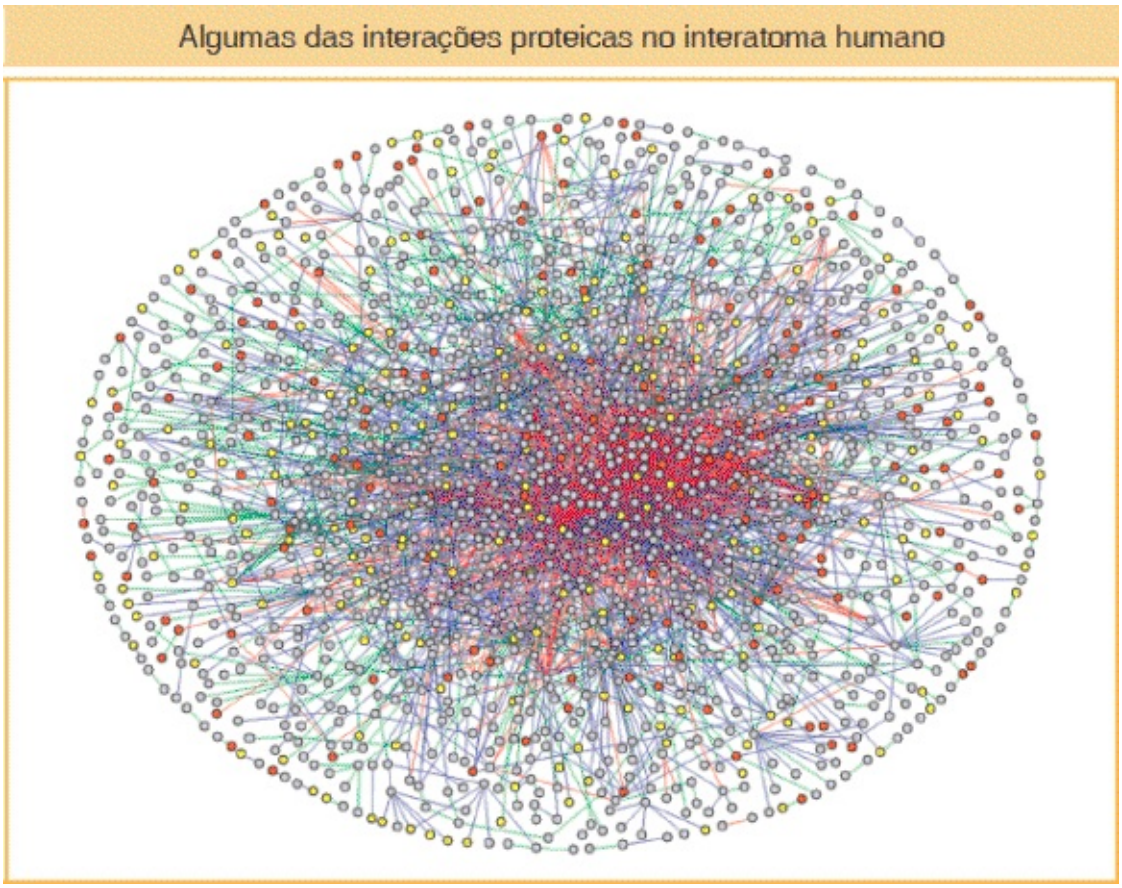


FIGURA 9.21 As proteínas (*representadas pelos círculos*) interagem com outras proteínas (*conectadas por linhas*) para formar complexos proteicos simples ou grandes. Este interatoma demonstra 3.186 interações de 1.705 proteínas humanas. (Dados de Ulrich Steizl *et al.*, Max Delbrück Center for Molecular Medicine [MDC] Berlin-Buch. Direitos autorais MDC.)

Qual é a significância biológica dessas interações? Neste capítulo e nos antecedentes, você observou que as interações de proteínas são centrais para a função de grandes máquinas biológicas tais como o replissomo, o spliceossomo e o ribossomo. Outro conjunto de interações significativas é a das associações entre as proteínas humanas e as proteínas dos patógenos humanos. Por exemplo, o interatoma de 40 proteínas do vírus Epstein-Barr (EBV) e de 112 proteínas humanas é composto por 173 interações (Figura 9.22). A compreensão sobre essa rede de interações pode levar a novas terapias para a mononucleose, uma doença causada por infecção pelo EBV.

Ubiquitinação. Surpreendentemente, uma das modificações pós-tradução mais

comuns não é sutil como a adição de um grupo fosfato. Em vez disso, essa modificação tem por alvo a degradação da proteína por meio de uma máquina biológica e da protease denominada proteossomo 26S (Figura 9.23). A modificação que marca uma proteína-alvo para ser degradada é a adição de cadeias de múltiplas cópias de uma proteína denominada **ubiquitina** aos resíduos de ε-amina da lisina (denominada **ubiquitinação**). A ubiquitina contém 76 aminoácidos e é observada apenas em eucariotos, nos quais é altamente conservada em plantas e em animais. Duas classes amplas de proteínas são alvo de destruição por meio da ubiquitinação: proteínas de vida curta, tais como reguladoras do ciclo celular, ou proteínas danificadas ou mutadas.

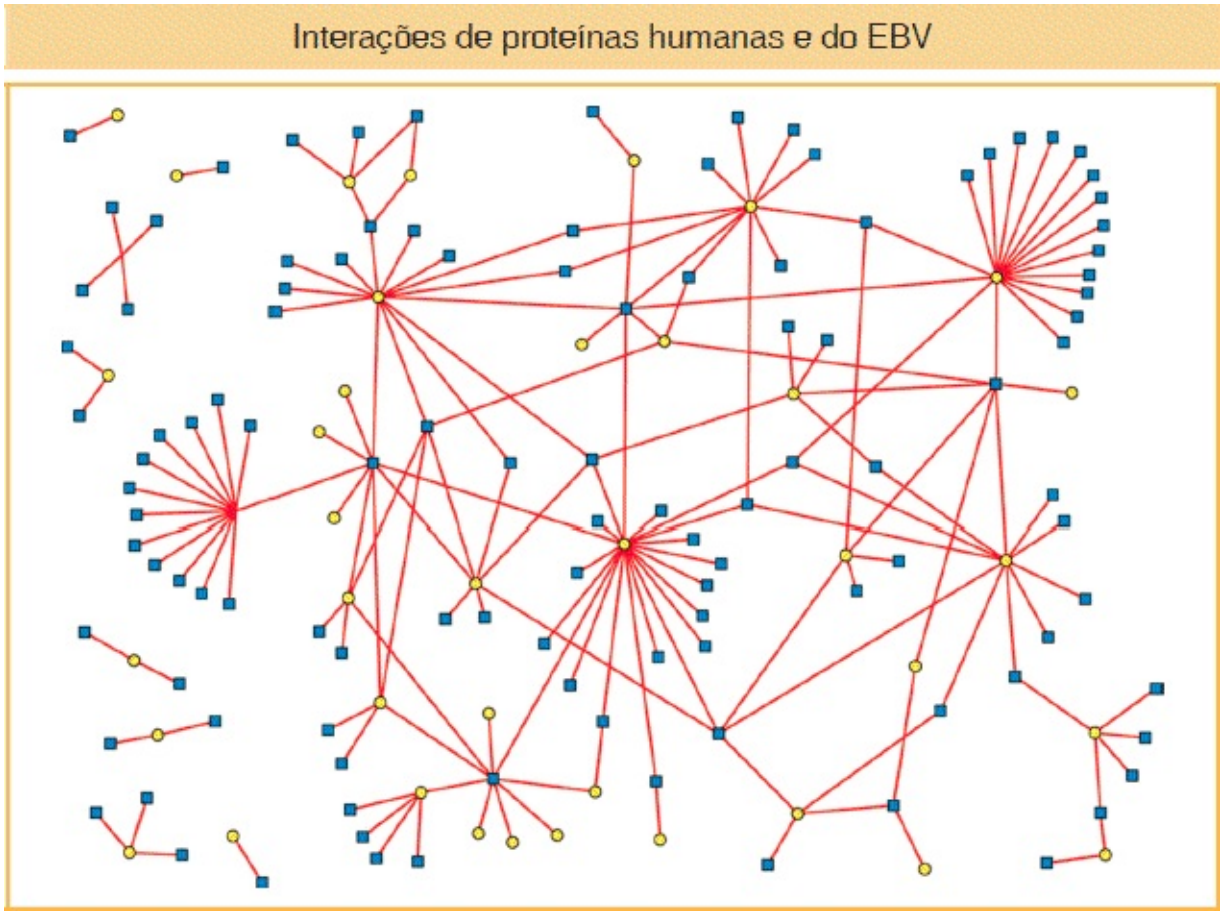


FIGURA 9.22 A rede de 173 interações de 40 proteínas do vírus Epstein-Barr (EBV) com 112 proteínas humanas. As proteínas virais estão demonstradas como *círculos amarelos* e as proteínas humanas como *quadrados azuis*. As interações estão demonstradas como *linhas vermelhas*. (Dados de Calderwood *et al.*, *Proceedings of the National Academy of Sciences* 104, 2007, 7606-7611. Direitos autorais 2007 por National

Academy of Sciences.)

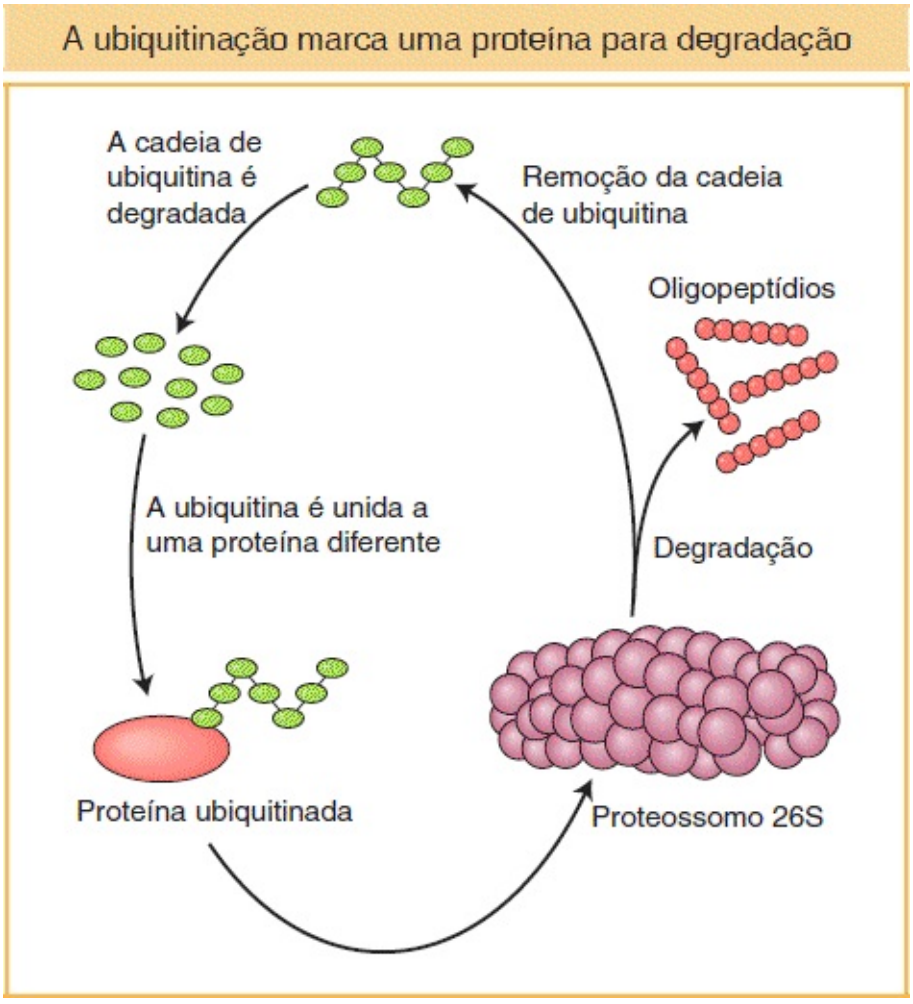


FIGURA 9.23 Estão demonstradas as principais etapas na degradação proteica mediada pela ubiquitina. A ubiquitina é primeiramente conjugada a outra proteína e em seguida degradada pelo proteossomo. Em seguida a ubiquitina e os oligopeptídios são reciclados.

Direcionamento de proteínas. Em eucariotos, todas as proteínas são sintetizadas nos ribossomos no citoplasma. Entretanto, algumas dessas proteínas acabam no núcleo, outras nas mitocôndrias e ainda outras ancoradas na membrana ou secretadas da célula. Como essas proteínas "sabem" onde elas devem ir? A resposta para esse problema aparentemente complexo de fato é consideravelmente simples: uma proteína recém-sintetizada contém uma sequência curta que a direciona para o local ou compartimento celular correto. Por exemplo, uma proteína de membrana recém-sintetizada ou uma proteína

destinada para uma organela apresenta um curto peptídio líder, denominado **sequência de sinal**, na sua extremidade amino-terminal. Para as proteínas de membrana, esse trecho de 15 a 25 aminoácidos direciona a proteína para canais na membrana do retículo endoplasmático, onde a sequência de sinal é clivada por uma peptidase (Figura 9.24). A partir do retículo endoplasmático, a proteína é direcionada para o seu destino final. Existe um fenômeno semelhante para determinadas proteínas bacterianas que são secretadas.

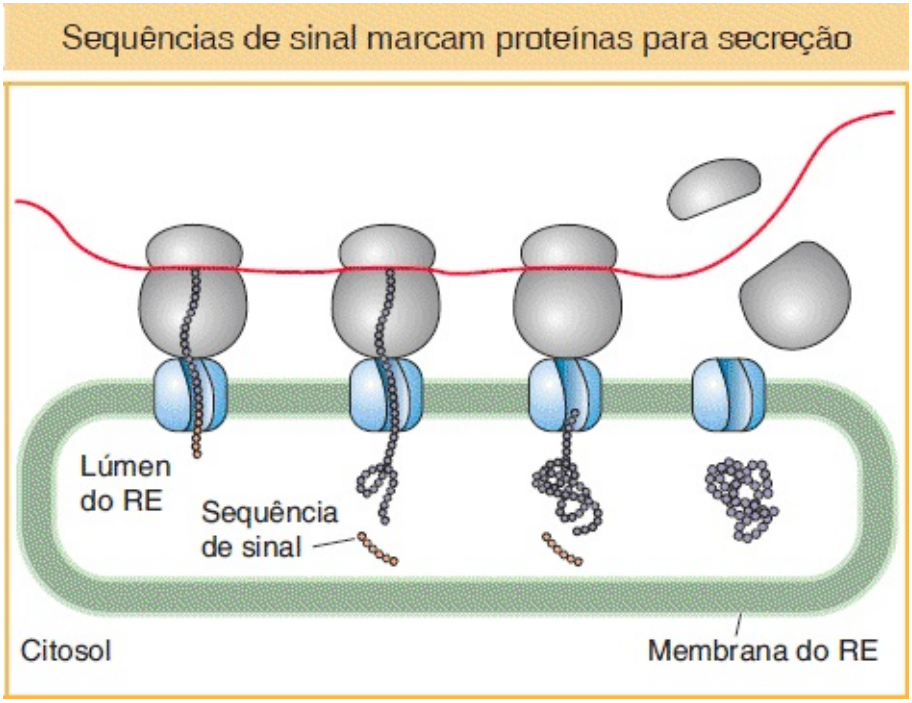


FIGURA 9.24 As proteínas destinadas a serem secretadas da célula apresentam uma sequência amino-terminal que é rica em resíduos hidrofóbicos. Essa sequência de sinal se liga às proteínas na membrana do retículo endoplasmático (RE) que retiram o restante da proteína por meio da bicamada lipídica. A sequência de sinal é clivada da proteína neste processo por uma enzima denominada *peptidase de sinal* (*não demonstrada*). Uma vez dentro do retículo endoplasmático, a proteína é direcionada para a membrana celular, a partir da qual será secretada.

As proteínas destinadas ao núcleo incluem as RNA e DNA polimerases e os fatores de transcrição discutidos nos Capítulos 7 e 8. As sequências de aminoácidos inseridas no interior das referidas proteínas ligadas ao núcleo são necessárias para o transporte do citoplasma para o núcleo. Essas **sequências de localização nuclear (NLS)** são reconhecidas por proteínas receptoras

citoplasmáticas que transportam proteínas recém-sintetizadas através dos poros nucleares — sítios na membrana por meio dos quais moléculas grandes são capazes de passar para dentro e para fora do núcleo. Uma proteína normalmente não observada no núcleo será direcionada para o núcleo se uma NLS estiver ligada a ela.

Por que as sequências de sinal são clivadas durante o direcionamento, enquanto uma NLS, localizada no interior de uma proteína, permanece após a proteína se movimentar para dentro do núcleo? Uma explicação pode ser que, na desintegração nuclear que acompanha a mitose (ver Capítulo 2), as proteínas localizadas no núcleo podem ser encontradas no citoplasma. Tendo em vista que essa proteína contém uma NLS, ela pode ser realocada para o núcleo de um célula-filha que resulta da mitose.

> CONCEITO-CHAVE A maioria das proteínas eucarióticas é inativa, exceto se modificada após a tradução. Alguns eventos pós-tradução, tais como a fosforilação ou a ubiquitinação, modificam os grupos laterais de aminoácidos, promovendo, assim, a ativação ou a degradação proteica, respectivamente. Outros mecanismos pós-tradução reconhecem a identidade dos aminoácidos em uma sequência proteica e direcionam aquelas proteínas para locais nos quais sua atividade é necessária dentro ou fora da célula.

RESUMO

Este capítulo descreveu a tradução da informação codificada na sequência de nucleotídios de um mRNA em uma sequência de aminoácidos de uma proteína. Nossas proteínas, mais do que qualquer outra macromolécula, determinam quem e o que somos. Elas são as enzimas responsáveis pelo metabolismo celular, incluindo a síntese de DNA e RNA, e são os fatores reguladores necessários para a expressão do programa genético. A versatilidade das proteínas como moléculas biológicas é manifestada na diversidade de formas que elas podem assumir. Além disso, até mesmo após a sua síntese, elas podem ser modificadas em uma

diversidade de modos por meio da adição de moléculas que conseguem alterar a sua função.

Tendo em vista o papel central das proteínas na vida, não é surpresa que ambos o código genético e o maquinário para a tradução desse código em proteínas tenham sido altamente conservados desde as bactérias até os seres humanos. Os principais componentes da tradução são três classes de RNA: tRNA, mRNA e rRNA. A precisão da tradução depende da ligação enzimática de um aminoácido com seu tRNA cognato, gerando uma molécula de tRNA carregada. Como adaptadores, os tRNA são as moléculas-chave na tradução. Contrariamente, o ribossomo é a fábrica na qual mRNA, tRNA carregados e outros fatores proteicos se unem para a síntese proteica.

A decisão-chave na tradução é onde iniciar a tradução. Em procariotos, o complexo de iniciação é montado no mRNA na sequência de Shine-Dalgarno, logo *upstream* do códon de início AUG. O complexo de iniciação em eucariotos é montado na estrutura *cap* 5′ do mRNA e se movimenta na direção 3′ até que o códon de início seja reconhecido. A fase mais longa da tradução é o ciclo de alongamento; nessa fase, o ribossomo se movimenta ao longo do mRNA, revelando o próximo códon que irá interagir com seu tRNA carregado cognato, de modo que os tRNA carregados com aminoácidos possam ser adicionados à cadeia de polipeptídios em crescimento. Esse ciclo continua até que um códon de parada seja encontrado. Os fatores de liberação facilitam o término da tradução.

Nos últimos anos, novas técnicas de imagem revelaram interações ribossômicas no nível atômico. Com esses novos “olhos”, atualmente podemos ver que o ribossomo é uma máquina incrivelmente dinâmica, que altera sua forma em resposta aos contatos realizados com tRNA e com proteínas. Além disso, as imagens em resolução atômica revelaram que os RNA ribossômicos, não as proteínas ribossômicas, estão intimamente associados aos centros funcionais do ribossomo.

O proteoma é o conjunto completo de proteínas que pode ser expresso pelo material genético de um organismo. Enquanto um eucarioto multicelular típico apresenta aproximadamente 20.000 genes, o proteoma típico provavelmente é 10 a 50 vezes maior. Essa diferença em parte é o resultado de modificações pós-

tradução, tais como fosforilação e ubiquitinação, que influenciam a atividade e a estabilidade proteica.

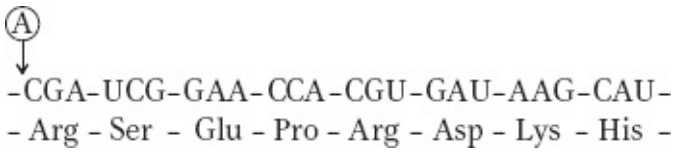
TERMOS-CHAVE

aminoácido
aminoacil-tRNA sintetase
anticódon
centro decodificador
centro peptidiltransferase
código degenerado
códon
desenho de fármaco com base na estrutura
direcionamento de proteínas
domínio
estrutura primária
estrutura quaternária
estrutura secundária
estrutura terciária
extremidade amino
extremidade carboxila
fator de iniciação
fator de liberação (RF)
iniciador
interatoma
isoforma
oscilante
polipeptídio
proteína fibrosa
proteína globular
proteoma
ribossomo

RNA ribossômico (rRNA)
RNA transportador ou de transferência (tRNA)
sequência de localização nuclear (NLS)
sequência de Shine-Dalgarno
sequência de sinal
sítio A
sítio ativo
sítio E
sítio P
subunidade
trinca
tRNA carregado
ubiquitina
ubiquitinação

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Com a utilização da Figura 9.5, demonstre as consequências sobre a tradução subsequente da adição de uma base adenina no início da sequência codificadora a seguir:



Solução

Com a adição da A no início da sequência codificadora, o quadro de leitura é alterado, e um conjunto diferente de aminoácidos é especificado pela sequência, conforme demonstrado aqui (observe que é encontrado um conjunto de códons sem sentido, que resulta no término da cadeia):

–ACG–AUC–GGA–ACC–ACG–UGA–UAA–GCA–
– Thr – Ile – Gly – Thr – parada – parada

Problema resolvido 2. A adição de um único nucleotídio seguida por uma deleção de um nucleotídio único à distância de aproximadamente 20 pb no DNA causa uma alteração na sequência proteica de:

–His–Thr–Glu–Asp–Trp–Leu–His–Gln–Asp–

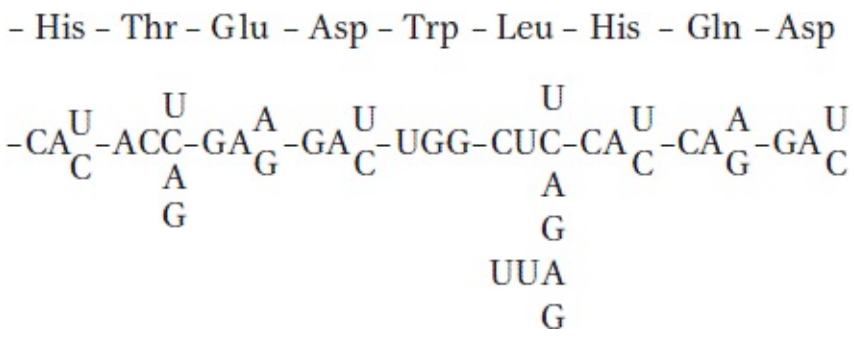
para:

–His–Asp–Arg–Gly–Leu–Ala–Thr–Ser–Asp–

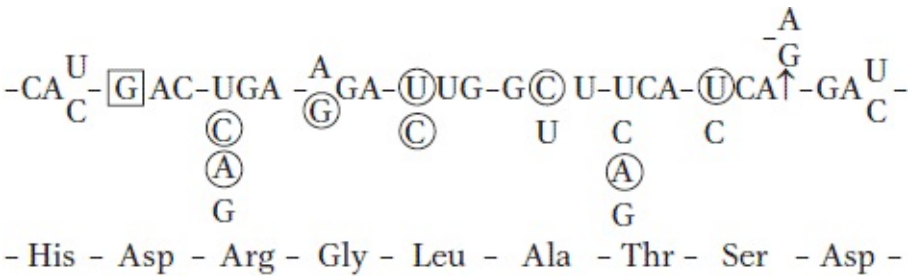
Qual nucleotídio foi adicionado e qual foi deletado? Qual é a sequência de mRNA original e a nova? (**Dica**: ver Figura 9.5).

Solução

Podemos desenhar a sequência de mRNA em relação à sequência proteica original (com as ambiguidades inerentes nesse estágio):



Tendo em vista que a alteração na sequência proteica fornecida a nós no início do problema tem início após o primeiro aminoácido (His) em virtude da adição de um nucleotídio único, podemos deduzir que um códon Thr tem de ser alterado para um códon Asp. Essa alteração necessariamente resulta da adição de uma G diretamente antes do códon Thr (indicado por um quadro), que altera o quadro de leitura, conforme demonstrado aqui:



Além disso, tendo em vista que uma deleção de um nucleotídio deve restaurar o códon Asp final para o quadro de leitura correto, uma A ou G deve ter sido deletada do final do códon próximo ao último original, conforme demonstrado pela seta. A sequência proteica original nos possibilita desenhar o mRNA com uma diversidade de ambiguidades. Entretanto, a sequência proteica que resulta do deslocamento do quadro de leitura nos possibilita determinar qual nucleotídio estava no mRNA original na maior parte desses pontos de ambiguidade. Os nucleotídios que podem ter aparecido na sequência original estão circulados. Em apenas alguns casos a ambiguidade permanece.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. A estrutura proteica primária está demonstrada na Figura 9.3 A. Onde no mRNA (próximo da extremidade 5′ ou 3′) uma mutação em R_2 seria codificada?

2. Neste capítulo, você conheceu mutações supressoras sem sentido nos genes do tRNA. Entretanto, as mutações supressoras também ocorrem em genes codificadores de proteínas. Com a utilização da estrutura terciária da subunidade β da hemoglobina demonstrada na Figura 9.3 C, explique em termos estruturais como uma mutação poderia causar a perda da função proteica da globina. Agora explique como uma mutação em um segundo local na mesma proteína poderia suprimir esta mutação e levar a uma proteína normal ou quase normal.

3. Com a utilização da estrutura quaternária da hemoglobina demonstrada na

Figura 9.3 D, explique em termos estruturais como uma mutação na proteína da subunidade β poderia ser suprimida por uma mutação no gene da subunidade α.

4. RNA transportadores (tRNA) são exemplos de moléculas de RNA que não codificam proteínas. Com base nas Figuras 9.6 e 9.8, qual é a significância da sequência de moléculas de tRNA? Qual você prevê que seria o impacto sobre a tradução de uma mutação em uma das bases de uma das hastes na estrutura do tRNA? No organismo mutante?

5. Os componentes de ribossomos procarióticos e eucarióticos estão demonstrados na Figura 9.10. Com base nesta figura, você acredita que o grande RNA ribossômico procariótico (rRNA 23S) seria capaz de substituir o rRNA 28S eucariótico? Justifique a sua resposta.

6. Os RNA ribossômicos (rRNA) são outro exemplo de uma molécula de RNA funcional. Com base na Figura 9.11, qual você acredita que seja a significância da estrutura secundária do rRNA?

7. Na Figura 9.12, o aminoácido terminal que está emergindo do ribossomo é codificado pela extremidade 5′ ou 3′ do mRNA?

8. Na Figura 9.12 B, o que você acredita que ocorra com o tRNA que é liberado do sítio E?

9. Na Figura 9.17, o que você acredita que ocorra próximo das subunidades ribossômicas após elas terminarem a tradução daquele mRNA?

10. Com base na Figura 9.19, você consegue prever a posição de uma mutação que afetaria a síntese de uma isoforma, mas não da outra?

11. Com base na Figura 9.24, você consegue prever a posição de uma mutação que produziria uma proteína ativa que não foi direcionada para o local correto?

PROBLEMAS BÁSICOS

12.

a. Utilize o dicionário de códons na Figura 9.5 para completar a tabela a

seguir. Presuma que a leitura seja da esquerda para a direita e que as colunas representem os alinhamentos da transcrição e da tradução.

C												Dupla- de DN
						T	G	A				
	C	A				U						mRNA transcr
									G	C	A	Anticó de tRN adequa
			Trp									Amino incorp à prote

b. Rotule as extremidades 5′ e 3′ do DNA e do RNA, bem como as extremidades amino e carboxila da proteína.

13. Considere o segmento de DNA a seguir:

5′ GCTTCCCAA 3′
3′ CGAAGGGTT 5′

Presuma que o filamento superior seja o filamento-molde utilizado pela RNA polimerase.

a. Desenhe o RNA transcrito.

b. Rotule as suas extremidades 5′ e 3′.

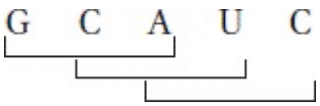
c. Desenhe a cadeia de aminoácidos correspondente.

d. Rotule as suas extremidades amino e carboxila.

Repita as partes *a* a *d*, presumindo que o filamento inferior seja o filamento-molde.

14. Um evento mutacional insere um par de nucleotídios extra no DNA. Qual dos desfechos a seguir você espera? (1) Absolutamente nenhuma proteína; (2) uma proteína na qual um aminoácido está alterado; (3) uma proteína na qual três aminoácidos estão alterados; (4) uma proteína na qual dois aminoácidos estão alterados; (5) uma proteína na qual a maior parte dos aminoácidos após o sítio da inserção está alterada.

15. Antes de a verdadeira natureza do processo do código genético ter sido totalmente compreendida, foi proposto que a mensagem poderia ser lida em trincas sobrepostas. Por exemplo, a sequência GCAUC poderia ser lida como GCA CAU AUC:



Planeje um teste experimental dessa ideia.

16. Se o tRNA é o adaptador para a tradução, o que é o ribossomo?

17. Qual anticódon você preveria para uma espécie de tRNA que carreia a isoleucina? Existe mais de uma resposta possível? Caso afirmativo, declare quaisquer respostas alternativas.

18. **a.** Em quantos casos no código genético você falharia em conhecer o aminoácido especificado por um códon se você soubesse apenas os primeiros dois nucleotídios do códon?

b. Em quantos casos você falharia em conhecer os primeiros dois nucleotídios do códon se soubesse qual aminoácido é especificado por ele?

19. Deduza quais seis códons do tipo selvagem podem ter estado nos mutantes que levaram Brenner a inferir a natureza do códon UAG.

20. Se um polirribonucleotídio contém quantidades iguais de bases adenina e

uracila posicionadas aleatoriamente, qual proporção de suas trincas irá codificar **(a)** fenilalanina, **(b)** isoleucina, **(c)** leucina, **(d)** tirosina?

21. Você sintetizou três RNA mensageiros diferentes com bases incorporadas em uma sequência aleatória nas razões a seguir: **(a)** 1 U: 5 C, **(b)** 1 A: 1 C: 4 U, **(c)** 1 A: 1 C: 1 G: 1 U. Em um sistema de síntese proteica *in vitro*, indique as identidades e as proporções de aminoácidos que serão incorporados nas proteínas quando cada um desses mRNA for testado. (Consulte a Figura 9.5.)

22. No fungo *Neurospora*, foram obtidos alguns mutantes que apresentavam ausência de atividade em relação a uma determinada enzima. Observou-se, por meio de mapeamento, que as mutações estão em um de dois genes não ligados. Forneça uma explicação possível em referência à estrutura proteica quaternária.

23. Qual é o significado da declaração "O código genético é universal"? Qual é a significância desse achado?

24. A enzima triptofano sintetase é produzida em dois tamanhos, grande e pequena. Alguns mutantes sem atividade enzimática produziram enzimas exatamente do mesmo tamanho que o tipo selvagem. Outros mutantes sem atividade produziram apenas a enzima grande; outros ainda, apenas a enzima pequena.

a. Explique os diferentes tipos de mutantes no nível da estrutura proteica.

b. Por que você acredita que não havia mutantes que não produziam nenhuma enzima?

25. Nos experimentos de Crick-Brenner descritos neste capítulo, três inserções ou três deleções restauraram o quadro de leitura normal e a dedução foi que o código era lido em grupos de três. Essa dedução realmente é comprovada pelos experimentos? Um códon pode ter sido composto por seis bases, por exemplo?

26. Um mutante não apresenta atividade em relação à enzima isocitrato liase. Como esse resultado comprova que a mutação está no gene que codifica a

isocitrato liase?

27. Um determinado supressor sem sentido corrige um mutante sem crescimento até um estado que é próximo, mas não exatamente, do tipo selvagem (ele apresenta crescimento anormal). Sugira um possível motivo pelo qual a reversão não é uma correção total.

28. Em genes bacterianos, assim que algum transcrito de mRNA parcial é produzido pelo sistema de RNA polimerase, o ribossomo se associa a ele e inicia a tradução. Desenhe um diagrama desse processo, identificando as extremidades 5′ e 3′ do mRNA, as extremidades COOH e NH da proteína, a RNA polimerase, e no mínimo um ribossomo. Por que esse sistema não poderia atuar em eucariotos?

29. Em um haploide, um supressor sem sentido *su1* atua na mutação 1, mas não na mutação 2 ou 3 do gene *P*. Um supressor sem sentido não ligado *su2* atua na mutação 2 de *P*, mas não sobre 1 ou 3. Explique esse padrão em relação à natureza das mutações e dos supressores.

30. Foram desenvolvidos sistemas de tradução *in vitro*, nos quais moléculas de RNA específicas podem ser adicionadas a um tubo de ensaio que contém um extrato celular bacteriano que inclui todos os componentes necessários para a tradução (ribossomos, tRNA, aminoácidos). Se um aminoácido marcado radioativamente for incluído, qualquer proteína traduzida a partir daquele RNA pode ser detectada e demonstrada em um gel. Se um mRNA eucariótico for adicionado ao tubo de ensaio, serão produzidas proteínas radioativas? Explique.

31. Um sistema de tradução *in vitro* contém um extrato de célula eucariótica que inclui todos os componentes necessários para a tradução (ribossomos, tRNA, aminoácidos). Se um RNA bacteriano for adicionado ao tubo de ensaio, uma proteína será produzida? Caso negativo, por que não?

32. Um sistema de tradução quimérico que contém a subunidade ribossômica grande de *E. coli* e a subunidade ribossômica pequena de levedura (um eucarioto unicelular) seria capaz de atuar na síntese proteica? Explique por

que sim ou por que não.

33. Mutações que alteram um único aminoácido no sítio ativo de uma enzima podem resultar na síntese de quantidades do tipo selvagem de uma enzima inativa. Você consegue pensar em outras regiões em uma proteína nas quais uma alteração de aminoácido único pode apresentar o mesmo resultado?

34. Qual evidência ampara a visão de que os RNA ribossômicos são um componente mais importante do ribossomo do que as proteínas ribossômicas?

35. Explique por que os antibióticos, tais como eritromicina e azitromicina, que se ligam à subunidade ribossômica grande não nos prejudicam.

36. Por que os eucariotos multicelulares precisam ter centenas de genes que codificam quinases?

37. Nosso sistema imune produz muitas proteínas diferentes que nos protegem contra infecções virais e bacterianas. Empresas de biotecnologia precisam produzir grandes quantidades dessas proteínas imunes para testes em humanos e a venda final para o público. Para essa finalidade, seus cientistas modificam culturas celulares bacterianas ou humanas para expressar essas proteínas imunes. Explique por que as proteínas isoladas de culturas bacterianas com frequência são inativas, enquanto as mesmas proteínas isoladas a partir das culturas celulares humanas são ativas (funcionais).

38. Você esperaria encontrar sequências de localização nuclear (NLS) nas proteínas que produzem DNA e RNA polimerases procarióticas e eucarióticas? Explique por que sim ou por que não.

PROBLEMAS DESAFIADORES

39. A adição de um nucleotídio único e a deleção de um nucleotídio único à distância de aproximadamente 15 bases no DNA causam uma alteração na sequência proteica de:

Phe–Ser–Pro–Arg–Leu–Asn–Ala–Val–Lys

para:

Phe–Val–His–Ala–Leu–Met–Ala–Val–Lys

a. Quais são as sequências de nucleotídios do mRNA antigo e do novo? (Utilize o dicionário de códons na Figura 9.5.)
b. Qual nucleotídio foi adicionado? Qual foi deletado?

40. Você está estudando um gene de *E. coli* que especifica uma proteína. Uma parte da sua sequência é:

–Ala–Pro–Trp–Ser–Glu–Lys–Cys–His–

Você recupera uma série de mutantes em relação a esse gene que não demonstram atividade enzimática. Ao isolar os produtos enzimáticos mutantes, você observa as sequências a seguir:

Mutante 1:

–Ala–Pro–Trp–Arg–Glu–Lys–Cys–His–

Mutante 2:

–Ala–Pro–

Mutante 3:

–Ala–Pro–Gly–Val–Lys–Asn–Cys–His–

Mutante 4:

–Ala–Pro–Trp–Phe–Phe–Thr–Cys–His–

Qual é a base molecular em relação a cada mutação? Qual é a sequência de DNA que especifica essa parte da proteína?

41. Atualmente são conhecidos supressores de mutações por deslocamento do quadro de leitura. Proponha um mecanismo para a sua ação.

42. Considere os genes que especificam a estrutura da hemoglobina. Disponha os eventos a seguir na sequência mais provável na qual eles ocorreriam:

a. É observada anemia.

b. A forma do sítio de ligação do oxigênio é alterada.

c. Um códon incorreto é transcrito no mRNA da hemoglobina.

d. O ovócito (gameta feminino) recebe uma dose alta de radiação.

e. Um códon incorreto é gerado no DNA de um gene de hemoglobina.

f. Uma mãe (uma técnica de radiologia) acidentalmente tropeça na frente de um gerador de raios X em operação.

g. Uma criança morre.

h. A capacidade de transporte de oxigênio do corpo está gravemente comprometida.

i. O anticódon do tRNA que se alinha é de um tipo que traz um aminoácido inadequado.

j. Ocorre a substituição de pares de nucleotídios no DNA de um gene de hemoglobina.

43. Quais características estruturais são compartilhadas pelos spliceossomos (Figuras 8.16 e 8.17) e os ribossomos? Por que ambas as estruturas são utilizadas para amparar a teoria do Mundo do RNA?

44. Uma molécula de DNA bifilamentar com a sequência demonstrada aqui produz, *in vivo*, um polipeptídio que apresenta o comprimento de cinco aminoácidos.

TACATGATCATTTCACGGAATTTCTAGCATGTA
ATGTACTAGTAAAGTGCCTTAAAGATCGTACAT

a. Qual filamento de DNA é o filamento-molde, e em qual direção ele é transcrito?

b. Rotule as extremidades 5′ e 3′ de cada filamento.

c. Se ocorre uma inversão entre a segunda e a terceira trincas das

extremidades esquerda e direita, respectivamente, e o mesmo filamento de DNA é transcrito, quão longo será o polipeptídio resultante?

d. Presuma que a molécula original esteja intacta e que o filamento inferior seja transcrito da esquerda para a direita. Forneça a sequência de bases do RNA e rotule as extremidades 5′ e 3′ do anticódon que insere o *quarto* aminoácido no polipeptídio nascente. Qual é o aminoácido?

45. Uma das técnicas que Khorana utilizou para decifrar o código genético foi sintetizar polipeptídios *in vitro*, utilizando mRNA sintético com diversas sequências de bases repetidas. Por exemplo, $(AGA)_n$, que pode ser escrita como AGAAGAAGAAGAAGA... Por vezes o polipeptídio resultante continha apenas um aminoácido (um homopolímero), e por vezes continha mais de um aminoácido (um heteropolímero), dependendo da sequência de repetição utilizada. Khorana observou que por vezes polipeptídios diferentes eram produzidos a partir do mesmo mRNA sintético, sugerindo que a iniciação da síntese proteica no sistema *in vitro* nem sempre tem início no primeiro nucleotídio do mensageiro. Por exemplo, a partir de $(CAA)_n$, três polipeptídios podem ter sido produzidos: homopolímero aa_1 (abreviado aa_1-aa_1), homopolímero aa_2 (aa_2-aa_2), e homopolímero aa_3 (aa_3-aa_3). Esses polipeptídios provavelmente correspondem às leituras a seguir derivadas do início em diferentes locais na sequência:

CAA CAA CAA CAA...
ACA ACA ACA ACA...
AAC AAC AAC AAC...

A tabela a seguir demonstra os resultados do experimento de Khorana.

mRNA sintético	Polipeptídio(s) sintetizado(s)
$(UG)_n$	(Ser-Leu)
$(UG)_n$	(Val-Cys)

$(AC)_n$	(Thr-His)
$(AG)_n$	(Arg-Glu)
$(UUC)_n$	(Ser-Ser) e (Leu-Leu) e (Phe-Phe)
$(UUG)_n$	(Leu-Leu) e (Val-Val) e (Cys-Cys)
$(AAG)_n$	(Arg-Arg) e (Lys-Lys) e (Glu-Glu)
$(CAA)_n$	(Thr-Thr) e (Asn-Asn) e (Gln-Gln)
$(UAC)_n$	(Thr-Thr) e (Leu-Leu) e (Tyr-Tyr)
$(AUC)_n$	(Ile-Ile) e (Ser-Ser) e (His-His)
$(GUA)_n$	(Ser-Ser) e (Val-Val)
$(GAU)_n$	(Asp-Asp) e (Met-Met)
$(UAUC)_n$	(Tyr-Leu-Ser-Ile)
$(UUAC)_n$	(Leu-Leu-Thr-Tyr)
$(GAUA)_n$	Nenhum
$(GUAA)_n$	Nenhum

Nota: a ordem na qual os polipeptídios ou aminoácidos estão listados na tabela não é significativa, exceto em relação a $(UAUC)_n$ e $(UUAC)_n$.

a. Por que $(GUA)_n$ e $(GAU)_n$ codificam, cada um, apenas dois homopolipeptídios?

b. Por que $(GAUA)_n$ e $(GUAA)_n$ falham em estimular a síntese?

c. Com a utilização dos resultados de Khorana, atribua um aminoácido para cada trinca na lista a seguir. Relembre que com frequência existem diversos códons para um único aminoácido e que as duas primeiras letras em um códon normalmente são importantes (mas que a terceira letra ocasionalmente é significativa). Também tenha em mente que alguns códons muito diferentes por vezes codificam o mesmo aminoácido. Tente solucionar este problema sem consultar a Figura 9.5.

GUA	GAU	UUG	AAC
GUG	UUC	UUA	GAA
GUU	AGU	UAU	AGA
AUG	CUU	AUC	GAG
UGU	CUA	UAC	CAA
ACA	UCU	AAG	UAG
CAC	CUC	ACU	UGA

A solução deste problema requer tanto lógica quanto tentativa e erro. Não se sinta desencorajado: Khorana recebeu um Prêmio Nobel por solucioná-lo. Boa sorte!

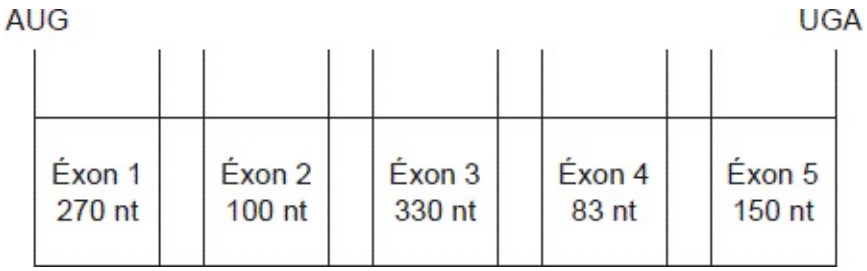
(Dados de J. Kuspira e G. W. Walker, *Genetics: Questions and Problems*. McGraw-Hill, 1973.)

46. Para construir um "interatoma" como aquele demonstrado na Figura 9.21, cientistas identificam todas as interações proteicas em um tipo de tecido ou célula em particular. A comparação dos interatomas do tecido muscular humano *versus* o tecido cerebral humano revela padrões muito diferentes. Se você fosse o cientista envolvido nesse estudo, como explicaria esses

resultados?

47. Os genomas da maioria dos eucariotos multicelulares codificam aproximadamente 25.000 genes, ainda que seus proteomas contenham mais de 200.000 proteínas. Proponha dois processos que, considerados em conjunto, expliquem essa discrepância.

48. Os éxons e os íntrons de um gene estão demonstrados a seguir. A recomposição alternativa desse gene produz três diferentes quadros de leitura abertos. Preveja quais éxons formarão esses três mRNA e forneça a justificativa para a sua resposta. nt = Nucleotídios



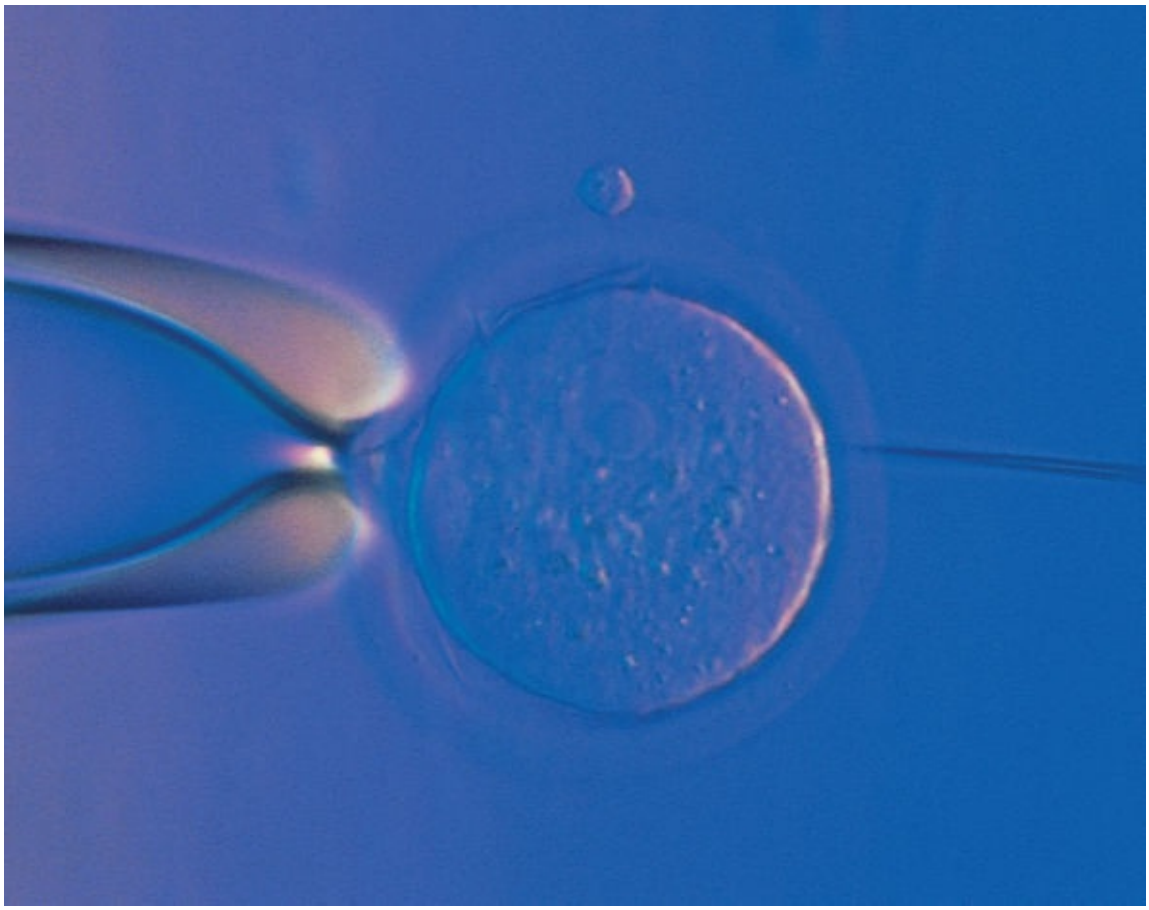
49. Se fosse encontrada vida em outro planeta, você acredita que ela apresentaria o mesmo código genético? Justifique sua resposta.

50. A imagem na primeira página do Capítulo 9 é de um ribossomo em resolução atômica, no qual o rRNA é azul e as proteínas ribossômicas são cor-de-rosa. Observe essa figura cuidadosamente, prestando atenção nos pontos de interação dos rRNA com as proteínas ribossômicas e das proteínas ribossômicas com os tRNA, mRNA e fatores ribossômicos (EF-Tu, RF1). Agora planeje um modelo que explique a natureza de algumas das interações (em termos muito gerais) que estão ocorrendo dentro dessa máquina molecular. Como o conhecimento sobre essas interações pode auxiliar no desenho de uma nova geração de antibióticos?

[1]F. Crick, “On Protein Synthesis.” *The Symposia of the Society for Experimental Biology* 12:138-163, 1958.

CAPÍTULO 10

Isolamento e Manipulação de Genes



Injeção de DNA exógeno em uma célula animal. A microagulha utilizada para a injeção está demonstrada *à direita*, e uma pipeta de manuseio celular está demonstrada *à esquerda*. (Rapho Agence/Science Source.)

TÓPICOS

10.1 Visão geral | Isolamento e amplificação de fragmentos específicos de DNA
10.2 Produção de moléculas de DNA recombinante

10.3 Utilização de sondas moleculares para encontrar e analisar um clone de interesse específico
10.4 Determinação da sequência de bases de um segmento de DNA
10.5 Alinhamento genético e mapas físicos para isolar genes específicos
10.6 Engenharia genética

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Diagramar as etapas por meio das quais um gene é isolado e amplificado por meio de clonagem
- Descrever quantos tipos diferentes de bibliotecas são utilizados para identificar moléculas de DNA específicas
- Comparar as técnicas utilizadas para amplificar o DNA com e sem clonagem
- Aplicar os diversos procedimentos utilizados para analisar o DNA isolado, o RNA isolado e as proteínas isoladas
- Contrastar as diferentes abordagens experimentais utilizadas para modificar os genomas de plantas e animais em laboratório
- Descrever as implicações das técnicas moleculares para a compreensão da função gênica.

Os genes são o foco central da genética e, assim, claramente seria desejável isolar um gene de interesse (ou alguma região do DNA) do genoma em uma quantidade adequada para estudo. O isolamento de genes individuais e a produção de uma quantidade suficiente deles para a análise pode ser uma tarefa desencorajadora, tendo em vista que um gene único é uma pequena fração de um genoma inteiro. Por exemplo, o genoma humano haploide contém mais de 3 bilhões de pares de bases, enquanto a região codificadora de um gene médio contém apenas alguns milhares de pares de bases. Como os cientistas encontram a agulha no palheiro, o gene, e em seguida produzem quantidades dele para análise?

Muitas investigações em genética têm início com o desejo de estudar um traço

ou uma doença. Com a utilização da genética direta, conforme descrito no Capítulo 2, procuramos mutantes que exibam um fenótipo alterado e, em seguida, realizamos cruzamentos ou analisamos heredogramas para determinar se aquele fenótipo é determinado por um gene único. No Capítulo 4, discutimos como o mapeamento por meio de recombinação auxilia na localização gênica no nível do DNA. Neste capítulo, continuamos com a apresentação dos métodos moleculares para a identificação de um gene de interesse e o estudo da sua função molecular.

A primeira etapa no estudo da função gênica é isolar o seu DNA e reproduzi-lo em quantidades adequadas para o estudo.

Assim como um operário da construção, um engenheiro genético necessita de ferramentas. A maior parte das caixas de ferramentas com as quais estamos familiarizados é preenchida com instrumentos como martelos, chaves de fenda e chaves inglesas, que são desenhadas por pessoas e manufaturadas em fábricas. Contrariamente, as ferramentas da engenharia genética são moléculas isoladas a partir de células. A maior parte dessas ferramentas foi produto de descobertas científicas — nas quais o objetivo era responder a uma questão biológica. Apenas posteriormente alguns cientistas reconheceram o possível valor prático de algumas dessas moléculas e inventaram modos para integrá-las em protocolos com o objetivo de isolar e amplificar fragmentos de DNA. Já fomos apresentados a algumas dessas moléculas nos capítulos anteriores, e neste capítulo você verá como elas se tornaram o fundamento da revolução biotecnológica.

Um modo de separar o nosso gene de interesse do restante do genoma é cortar o genoma com uma "tesoura molecular" e isolar o pequeno fragmento que contém o gene. Werner Arber descobriu essas tesouras moleculares, e por sua descoberta recebeu o Prêmio Nobel de Fisiologia ou Medicina em 1978. Entretanto, Arber não estava procurando uma ferramenta para cortar o DNA precisamente. Em vez disso, ele estava tentando compreender o motivo de algumas bactérias serem resistentes à infecção por vírus bacterianos. Ao responder a essa questão biológica, ele descobriu que bactérias resistentes possuem uma enzima anteriormente desconhecida — uma endonuclease de restrição — que corta o DNA em sequências específicas. A enzima que ele descobriu, *Eco*RI, tornou-se a primeira tesoura molecular comercialmente disponível.

Como outro exemplo, é improvável que alguém houvesse previsto que a DNA polimerase, enzima descoberta por Arthur Kornberg, uma descoberta pela qual ele recebeu o Prêmio Nobel de Fisiologia ou Medicina em 1959, pudesse transformar-se em outra ferramenta poderosa para o isolamento e a análise do DNA (ver Capítulo 7). Até hoje, muitas das técnicas utilizadas para determinar a sequência de nucleotídios do DNA depende da síntese de DNA com a DNA polimerase. De modo semelhante, a maior parte dos protocolos utilizados para isolar e amplificar regiões específicas do DNA de fontes tão díspares quanto uma cena de crime até um fóssil inserido em âmbar depende da atividade da DNA polimerase.

Tecnologia do DNA é um termo que descreve as técnicas coletivas para a obtenção, a amplificação e a manipulação de fragmentos de DNA específicos. Desde meados da década de 1970, o desenvolvimento da tecnologia do DNA revolucionou o estudo da biologia, abrindo muitas áreas de pesquisa para a investigação molecular. A **engenharia genética**, a aplicação da tecnologia do DNA para problemas biológicos, médicos, ou agrícolas específicos, agora é um ramo bem-estabelecido da tecnologia. **Genômica** é a extensão final da tecnologia para a análise global dos ácidos nucleicos presentes em um núcleo, uma célula, um organismo ou um grupo de espécies correlatas (ver Capítulo 14). Posteriormente neste capítulo, veremos como a tecnologia do DNA e a genômica, juntamente com os métodos apresentados nos Capítulos 2 e 4, podem ser utilizados em conjunto para isolar e identificar um gene.

10.1 Visão geral | Isolamento e amplificação de fragmentos específicos de DNA

Como um segmento de DNA específico pode ser isolado de um genoma inteiro? Além disso, como ele pode ser isolado em quantidades suficientes para analisar as características do DNA, tais como a sua sequência de nucleotídios e o seu produto proteico? Uma percepção crucial para a solução desse problema foi que os pesquisadores poderiam utilizar o maquinário de replicação do DNA (ver

Capítulo 7) para replicar o segmento de DNA em questão. A referida replicação é denominada **amplificação**. Ela pode ser realizada em células bacterianas vivas (*in vivo*) ou em um tubo de ensaio (*in vitro*).

Na abordagem *in vivo* (Figura 10.1 A), um investigador inicia com uma amostra de moléculas de DNA que contêm o gene de interesse. Essa amostra é denominada **DNA doador** e, na maioria das vezes, é um genoma inteiro. Fragmentos do DNA doador são inseridos em um plasmídio ou vírus bacteriano especialmente projetado, que "transportará" e amplificará o gene de interesse e que, portanto, são denominados **vetores**. Primeiramente, as moléculas do DNA doador são cortadas por meio da utilização de enzimas denominadas endonucleases de restrição como "tesouras" moleculares. Elas cortam longas moléculas de DNA do tamanho de cromossomos em centenas ou milhares de fragmentos de tamanho mais manuseável. Em seguida, cada fragmento é inserido em um cromossomo vetor que foi cortado para formar moléculas de **DNA recombinante**. As moléculas de DNA recombinante são transferidas para células bacterianas e, em geral, apenas uma molécula recombinante é captada por célula. Dentro da célula bacteriana, a molécula recombinante é amplificada juntamente com o vetor durante a divisão celular. A partir de uma única célula, esse processo resulta em um *clone* de células idênticas, cada um contendo a molécula de DNA recombinante e, assim, essa técnica de amplificação é denominada **clonagem de DNA**. Tendo em vista que muitos fragmentos de DNA são inseridos no vetor, a mistura de células resultante contém coletivamente tanto quanto o genoma inteiro do organismo doador. O próximo estágio é encontrar o clone raro que contém o DNA de interesse entre as muitas células.

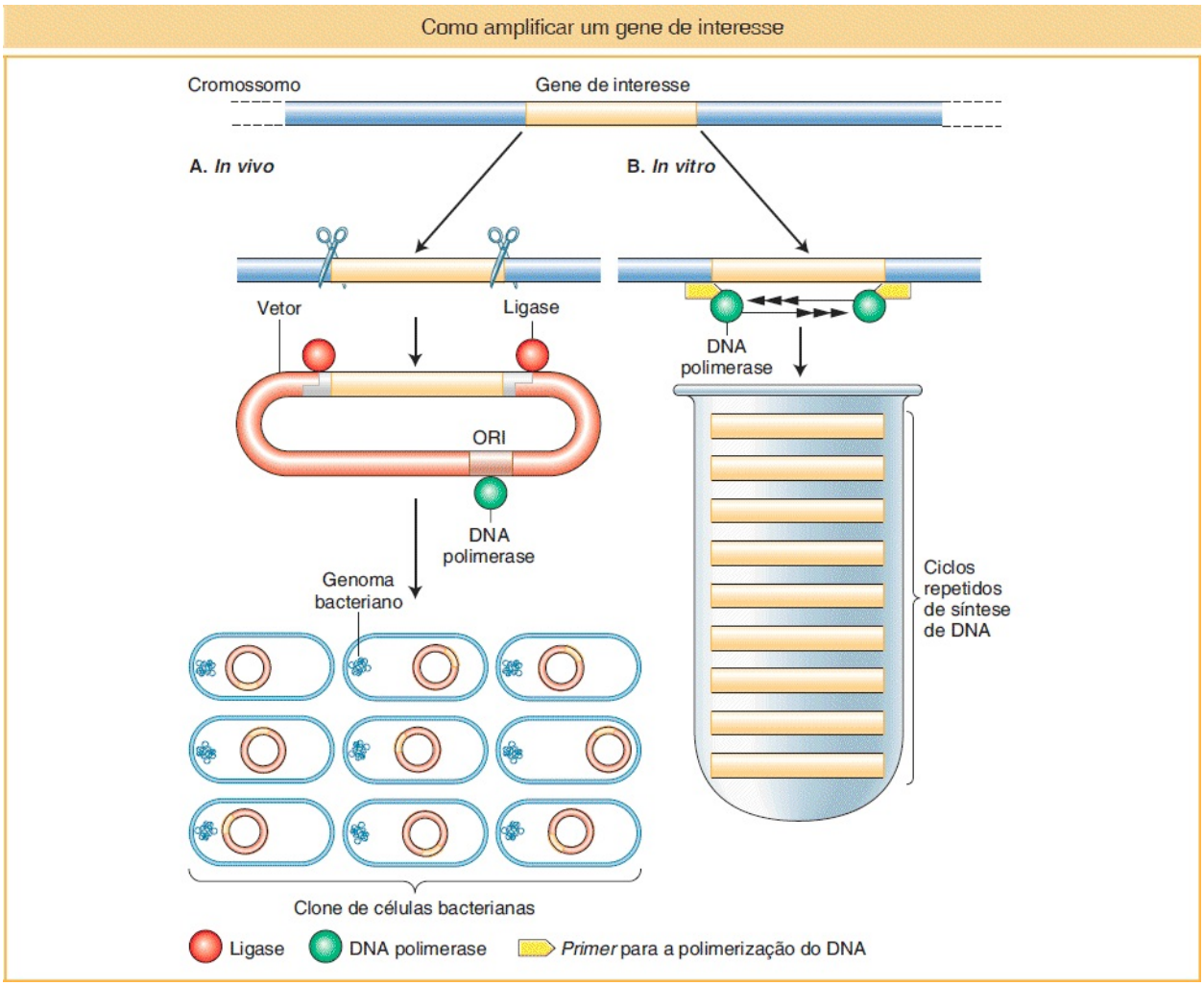


FIGURA 10.1 Dois métodos de isolamento e amplificação de um gene são (**A**) *in vivo*, enganando-se o maquinário de replicação de uma bactéria para que amplifique o DNA recombinante que contém o gene, e (**B**) *in vitro*, em tubo de ensaio, com a utilização da técnica de reação da cadeia de polimerase. Ambos os métodos empregam os princípios básicos da biologia molecular: a capacidade de proteínas específicas (*vermelhas* e *verdes*) de se ligarem ao DNA e a capacidade de segmentos de ácidos nucleicos unifilamentares complementares hibridizarem (o *primer* utilizado no método do tubo de ensaio).

Na abordagem *in vitro*, denominada **reação da cadeia de polimerase (PCR)** (Figura 10.1 B), um gene específico ou uma região do DNA de interesse são isolados e amplificados pela DNA polimerase. A PCR "encontra" a região do DNA de interesse (denominada DNA-alvo) por meio da ligação complementar de moléculas de DNA sintéticas curtas específicas, denominadas *primers*, às extremidades daquela sequência. Esses *primers* em seguida guiam o processo de replicação direcionado pela DNA polimerase, que ocorre em ciclos

exponenciais, resultando na produção de grandes quantidades do DNA-alvo como um fragmento de DNA isolado. Quantidades ainda maiores de DNA-alvo podem ser obtidas por meio da inserção do produto da PCR em um plasmídio (ou seja, o método *in vivo*), gerando, portanto, uma molécula de DNA recombinante como aquela descrita anteriormente.

Veremos repetidamente que a tecnologia do DNA depende de dois fundamentos básicos da pesquisa em biologia molecular:

- A capacidade de proteínas específicas reconhecerem e se ligarem a sequências de bases específicas dentro da dupla-hélice do DNA (os exemplos estão demonstrados em *verde* e em *vermelho* na Figura 10.1)
- A capacidade de sequências unifilamentares de DNA ou de RNA complementares parearem para formar moléculas bifilamentares (um exemplo é a ligação dos *primers* demonstrada em amarelo na Figura 10.1).

O restante do capítulo explorará algumas aplicações do DNA amplificado. Essas aplicações variam desde o isolamento gênico de rotina para pesquisas biológicas básicas, passando pela terapia gênica para tratar doenças humanas, e chegando até a produção de herbicidas e pesticidas para plantações. Para ilustrar como o DNA recombinante é produzido, consideraremos a clonagem do gene da insulina humana, um hormônio proteico utilizado no tratamento do diabetes. O diabetes melito é uma doença na qual os níveis de açúcar no sangue estão anormalmente altos, seja em virtude de o corpo não produzir insulina suficiente (diabetes tipo I) ou de as células serem incapazes de responder à insulina (diabetes tipo II). Formas brandas de diabetes tipo I podem ser tratadas por meio de restrições alimentares, mas, para muitos pacientes, são necessários tratamentos diários com insulina. Até aproximadamente 30 anos atrás, as vacas eram a principal fonte da proteína insulina, que era obtida a partir do pâncreas de animais abatidos em frigoríficos e purificada em grande escala para eliminar a maioria das proteínas e outros contaminantes dos extratos de pâncreas. Em seguida, em 1982, a primeira insulina humana recombinante chegou ao mercado. A insulina humana pode ser produzida em uma forma mais pura, a um custo mais baixo e em escala industrial,

por ser produzida em maiores quantidades em bactérias por meio de técnicas do DNA recombinante que utilizam a sequência real do gene da insulina humana. Além disso, não há o risco de introdução de vírus bovinos ou de uma resposta imune contra a insulina bovina. Utilizaremos o isolamento e a produção da insulina recombinante como um exemplo das etapas gerais necessárias para a produção de qualquer DNA recombinante.

10.2 Produção de moléculas de DNA recombinante

As moléculas de DNA recombinante normalmente contêm um fragmento de DNA inserido em um vetor bacteriano. Nesta seção, você verá que existem muitos tipos de moléculas de DNA recombinante que podem ser construídas a partir de uma diversidade de DNA doadores e de vetores. Iniciamos discutindo as fontes de DNA doador:

- Se o pesquisador deseja obter uma coleção de insertos que represente o genoma inteiro de um organismo, o DNA genômico pode ser cortado antes da clonagem
- Alternativamente, se o objetivo é isolar um gene único, a reação da cadeia de polimerase pode ser utilizada para amplificar regiões selecionadas do DNA *in vitro*
- Finalmente, se o pesquisador deseja apenas as sequências codificadoras de genes, sem íntrons, cópias de DNA dos produtos de mRNA, denominadas cDNA, podem ser sintetizadas e inseridas em um vetor.

O DNA genômico pode ser cortado antes da clonagem

O DNA genômico é obtido diretamente a partir de cromossomos do organismo em estudo, normalmente por meio da trituração do tecido fresco e da purificação do DNA. O DNA cromossômico pode ser utilizado como o ponto de início para ambos os métodos, *in vivo* e de PCR, a fim de isolar genes. Em relação ao método *in vivo*, o DNA genômico precisa ser cortado para que a clonagem seja possível. Conforme descrito posteriormente nesta seção, o DNA genômico não precisa ser cortado para a realização da PCR, tendo em vista que os *primers*

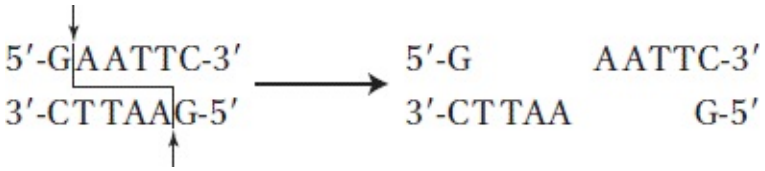
curtos específicos se ligam a ele e identificam o sítio de início para a DNA polimerase que direciona a replicação do DNA interveniente

As longas moléculas de DNA do tamanho de um cromossomo do DNA genômico devem ser cortadas em fragmentos de tamanho muito menor antes que possam ser inseridas em um vetor. A maior parte dos cortes é realizada com a utilização de **enzimas de restrição** bacterianas. Essas enzimas realizam cortes em sequências de DNA específicas, denominadas *sítios de restrição*, e essa propriedade é uma das características-chave que tornam as enzimas de restrição adequadas para a manipulação do DNA. Essas enzimas são exemplos de endonucleases, que clivam uma ligação fosfodiéster entre nucleotídios no DNA. Puramente ao acaso, qualquer molécula de DNA, de qualquer organismo, pode conter sítios de reconhecimento de enzima de restrição. Portanto, uma enzima de restrição cortará o DNA em um conjunto de **fragmentos de restrição** determinados pelos locais dos sítios de restrição, e produzirá o mesmo padrão de fragmentos a cada vez que ele for cortado.

Outra propriedade-chave de algumas enzimas de restrição é que muitas criam "extremidades coesivas" nos fragmentos. Vejamos um exemplo. A enzima de restrição *Eco*RI (da *E. coli*) reconhece a sequência a seguir de seis pares de nucleotídios no DNA de qualquer organismo:

5′-GAATTC-3′
3′-CTTAAG-5′

Esse tipo de segmento é denominado **palíndromo de DNA**, o que significa que ambos os filamentos apresentam a mesma sequência de nucleotídios, porém em orientação antiparalela (a leitura de 5′ para 3′ produz a mesma sequência em qualquer filamento). Diferentes enzimas de restrição realizam cortes em diferentes sequências palindrômicas. Por vezes, os cortes são na mesma posição em cada um dos dois filamentos antiparalelos, deixando *extremidades cegas*. Entretanto, a maior parte das enzimas de restrição úteis realiza cortes que são alternados ou em zigue-zague. A enzima *Eco*RI realiza cortes apenas entre os nucleotídios G e A em cada filamento do palíndromo:



Esses cortes alternados deixam um par de extremidades, e cada um deles apresenta uma extremidade unifilamentar de quatro bases (AATT). As extremidades são denominadas *coesivas* em virtude de, sendo unifilamentares, poderem realizar pareamento de bases (ou seja, aderir) com uma sequência complementar. O pareamento de filamentos únicos complementares desse tipo é denominado **hibridização**. A Figura 10.2 ilustra a enzima de restrição *Eco*RI realizando cortes alternados, em zigue-zague, nos dois filamentos em uma molécula de DNA circular, tal como um plasmídio; o corte abre o círculo, e a molécula linear resultante apresenta duas extremidades coesivas. Ela agora pode hibridizar com um fragmento de uma molécula de DNA diferente que apresenta as mesmas extremidades coesivas complementares.

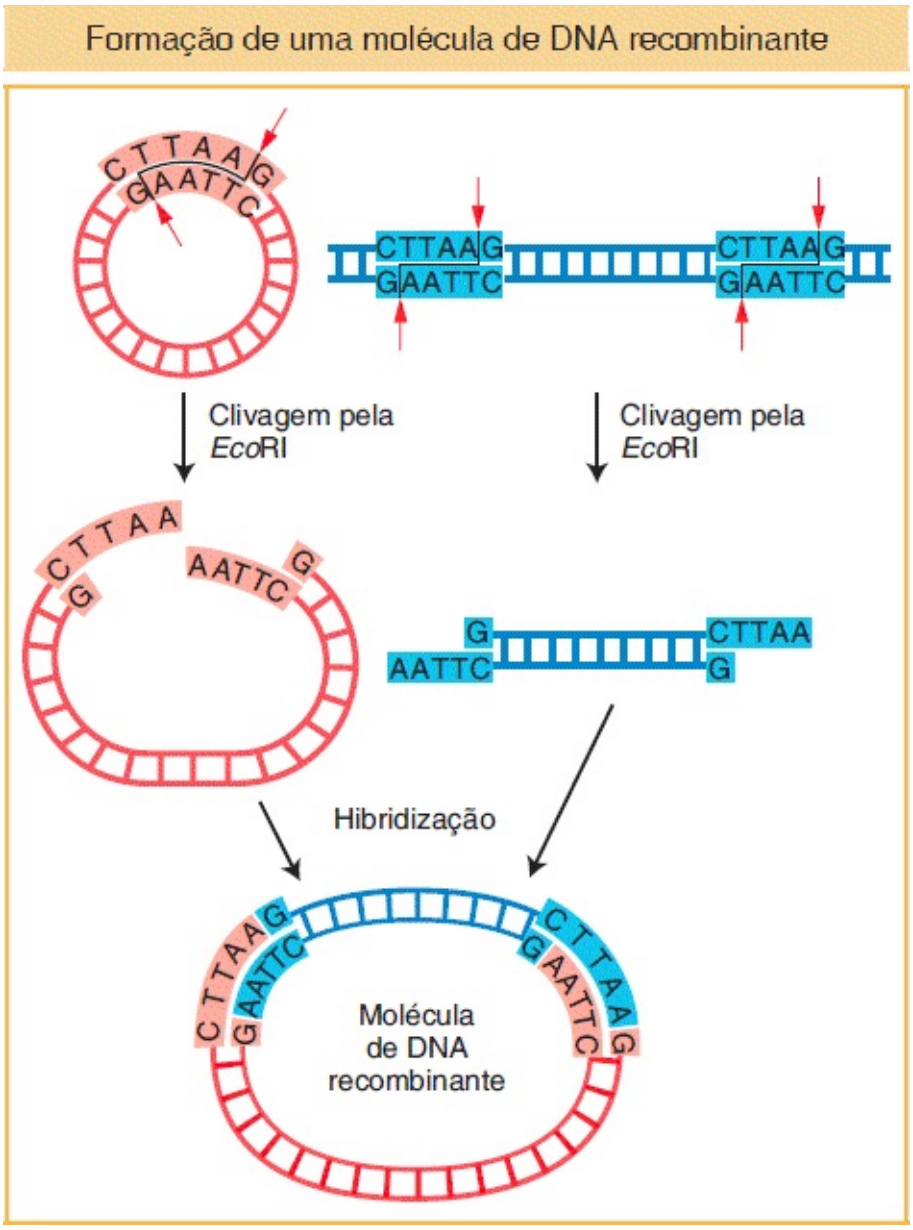


FIGURA 10.2 Para formar uma molécula de DNA recombinante, a enzima de restrição *Eco*RI corta uma molécula de DNA circular que contém uma sequência-alvo, resultando em uma molécula linear com extremidades unifilamentares coesivas. Em virtude da complementaridade, outras moléculas lineares com extremidades coesivas cortadas pela *Eco*RI podem hibridizar com o DNA circular linearizado, formando uma molécula de DNA recombinante.

A digestão do DNA genômico humano com *Eco*RI gera aproximadamente 500.000 fragmentos. Você verá posteriormente nesta seção como os cientistas analisaram esses fragmentos para encontrar a agulha no palheiro — um ou dois fragmentos dentre os 500.000 que contêm a sequência de DNA de interesse (em

nosso exemplo, o gene da insulina humana).

> CONCEITO-CHAVE O DNA genômico pode ser utilizado diretamente para a clonagem de genes. Como primeira etapa, as enzimas de restrição cortam o DNA em fragmentos de tamanho manuseável, e muitas delas geram extremidades coesivas unifilamentares adequadas para a produção do DNA recombinante.

A reação da cadeia de polimerase amplifica regiões selecionadas do DNA in vitro

Se nos esforçássemos para clonar o gene da insulina humana hoje, armados com a sequência genômica humana, o conhecimento do gene e das sequências flanqueadoras nos possibilitaria utilizar um método mais direto. Atualmente, podemos simplesmente amplificar o gene *in vitro* por meio da utilização da *reação da cadeia de polimerase (PCR)*. A estratégia básica da PCR está resumida na Figura 10.3. O processo utiliza múltiplas cópias de um par de *primers* de DNA curtos quimicamente sintetizados, com comprimento de aproximadamente 20 bases, desenhados de modo que cada *primer* se ligue a uma extremidade do gene ou da região a ser amplificada. Os dois *primers* se ligam a filamentos de DNA opostos que circundam a sequência-alvo, com suas extremidades 3′ apontando uma em direção à outra. As DNA polimerases adicionam bases às extremidades 3′ desses *primers* e copiam a sequência-alvo. A repetição do processo de polimerização produz um número exponencialmente crescente de moléculas de DNA bifilamentares. Os detalhes são como segue.

Iniciamos com uma solução que contém a fonte de DNA, os *primers*, os quatro desoxirribonucleotídios trifosfato (necessários para a síntese de DNA; ver Figura 7.15) e uma DNA polimerase tolerante ao calor. O DNA-alvo é desnaturado pelo calor (95°C), resultando em moléculas de DNA unifilamentares. Quando a solução é resfriada (entre 50 e 65°C), os *primers* hibridizam (ou *pareiam*) com suas sequências complementares nas moléculas de DNA unifilamentares. Após a

temperatura ser elevada até 72°C, a DNA polimerase termoestável replica os segmentos unifilamentares de DNA que se estendem a partir de um *primer.* A DNA polimerase *Taq*, da bactéria *Thermus aquaticus*, é uma enzima comumente utilizada. (Para sobreviver no calor extremo das fontes hidrotermais, essa bactéria desenvolveu proteínas que são extremamente resistentes ao calor. Portanto, sua DNA polimerase sobrevive a altas temperaturas necessárias para desnaturar a dupla-hélice de DNA, que desnaturariam e inativariam a DNA polimerase da maior parte das outras espécies.) Novos filamentos complementares são sintetizados assim como na replicação normal do DNA nas células, formando duas moléculas bifilamentares de DNA idênticas à molécula bifilamentar parental. Um ciclo é composto por essas três etapas, levando a uma replicação única do segmento entre os dois *primers*.

Reação da cadeia de polimerase

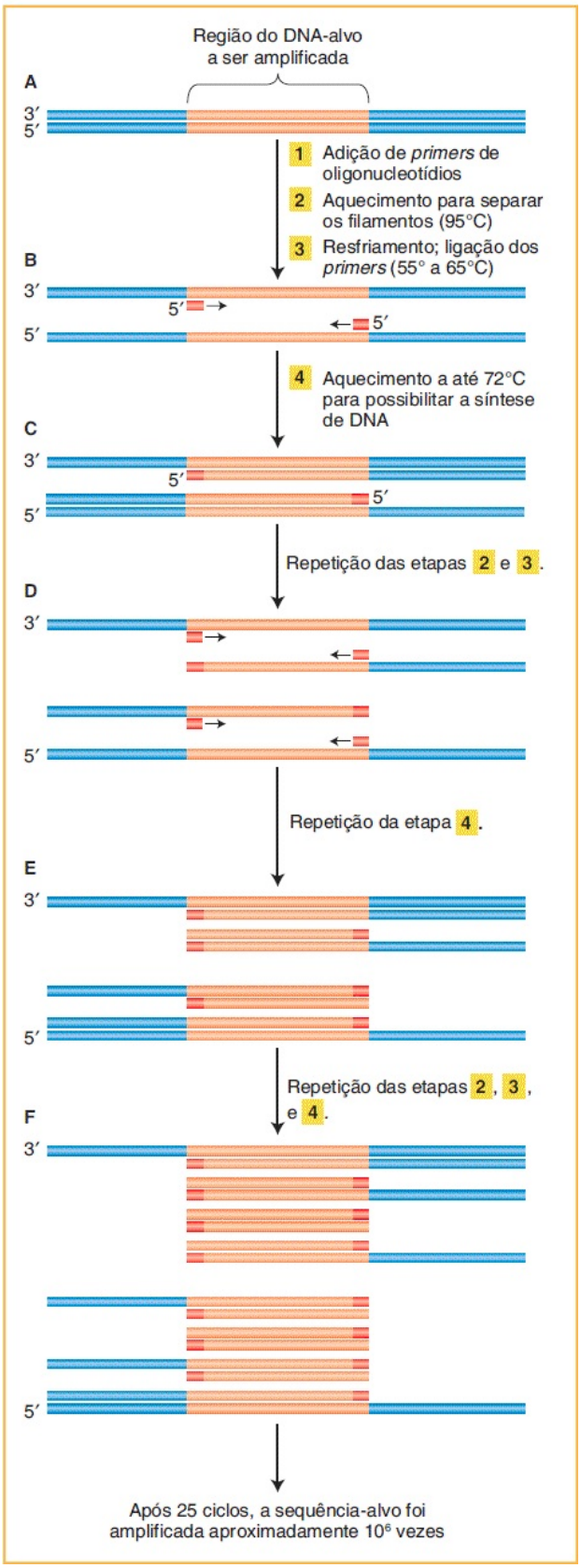


FIGURA 10.3 A reação da cadeia de polimerase copia rapidamente uma sequência de DNA-alvo. **A.** DNA bifilamentar que contém a sequência-alvo. **B.** Dois *primers* escolhidos ou sintetizados apresentam sequências complementares aos sítios de ligação do *primer* nas extremidades 3′ do gene-alvo nos dois filamentos. Os filamentos são separados por meio de aquecimento, em seguida são resfriados para possibilitar que os dois *primers* pareiem com os sítios de ligação de *primers*. Em conjunto, os *primers* assim flanqueiam a sequência-alvo. **C.** Após a elevação da temperatura, a polimerase *Taq* em seguida sintetiza o primeiro conjunto de filamentos complementares por meio da adição dos quatro nucleotídios trifosfato que também estão na mistura da reação. Estes dois primeiros filamentos são de comprimento variável, tendo em vista que não apresentam um sinal de parada comum. Eles se estendem além das extremidades da sequência-alvo, conforme delineado pelos sítios de ligação de *primers*. **D.** Os dois duplexes são novamente aquecidos, expondo quatro sítios de ligação. Após o resfriamento, os dois *primers* pareiam novamente com seus respectivos filamentos nas extremidades 3′ da região-alvo. **E.** Após a elevação da temperatura, a polimerase *Taq* sintetiza quatro filamentos complementares. Embora os filamentos-molde neste estágio sejam de comprimento variável, dois dos quatro filamentos recém-sintetizados a partir deles apresentam precisamente o comprimento da sequência-alvo desejada. Esse comprimento preciso é conquistado em virtude de cada um desses filamentos ter início no sítio de ligação do *primer*, em uma extremidade da sequência-alvo, e prosseguir até sair do molde, na outra extremidade da sequência. **F.** O processo é repetido por muitos ciclos, a cada vez criando mais moléculas bifilamentares de DNA idênticas à sequência-alvo.

Após a replicação do segmento entre os dois *primers* ser concluída, os dois novos duplexes são desnaturados pelo calor mais uma vez, para gerar moldes unifilamentares, e um segundo ciclo de replicação é realizado por meio da redução da temperatura na presença de todos os componentes necessários para que a polimerização produza quatro duplexes idênticos. Ciclos repetidos de desnaturação, pareamento e síntese resultam em um aumento exponencial no número de segmentos replicados. Tendo em vista que um ciclo típico dura 5 minutos, a amplificação até um bilhão de vezes pode ser prontamente realizada em 2,5 h. Conforme você verá posteriormente nesta seção, os produtos da PCR podem ser adicionalmente amplificados por meio de sua clonagem em células bacterianas.

A PCR é uma técnica poderosa que é utilizada rotineiramente para isolar genes específicos ou fragmentos de DNA quando existe um conhecimento anterior a respeito da sequência a ser amplificada. De fato, se todas as sequências correspondentes aos *primers* estiverem presentes apenas uma vez no genoma e estiverem suficientemente próximas, o *único* segmento de DNA que pode ser amplificado é aquele entre os dois *primers*. A PCR é uma técnica muito sensível,

com diversas aplicações em biologia. Ela pode amplificar sequências-alvo que estão presentes em números de cópias extremamente baixos em uma amostra, desde que sejam utilizados *primers* específicos para essa sequência rara. Por exemplo, investigadores criminais podem amplificar segmentos de DNA humano a partir das poucas células foliculares que circundam um único fio de cabelo retirado. Se os investigadores optarem por realizar isso, eles podem amplificar o gene da insulina a partir dessa amostra de DNA utilizando a sua localização precisa no cromossomo 11 para desenhar *primers* flanqueadores para a PCR direta.

Não seria exagerado dizer que a PCR revolucionou o estudo de muitos campos da biologia nos quais é necessária a análise do DNA. Em reconhecimento de sua importância para a ciência, Kary Mullis recebeu o Prêmio Nobel de Química em 1993 pelo desenvolvimento do primeiro protocolo de PCR viável.

> CONCEITO-CHAVE A reação da cadeia de polimerase utiliza *primers* especialmente desenhados para o isolamento direto e a amplificação de regiões específicas do DNA em um tubo de ensaio.

Cópias do DNA podem ser sintetizadas a partir de mRNA

Conforme vimos no Capítulo 8, os genes eucarióticos com frequência contêm um ou mais íntrons que rompem as regiões codificadoras. Além disso, conforme veremos nos Capítulos 14 e 15, os genes codificadores de proteínas com frequência correspondem a menos de 5% do DNA genômico de eucariotos multicelulares. Conforme mencionado na seção anterior, o gene da insulina humana contém dois íntrons, um problema se o objetivo for criar bactérias que sintetizem a insulina humana, tendo em vista que as bactérias não apresentam a capacidade de remover os íntrons presentes no DNA genômico natural. Em vez disso, tendo em vista que estamos interessados apenas na sequência codificadora, podemos utilizar o mRNA da insulina como material de início para a PCR. Em relação à insulina e outros genes codificadores de proteínas em eucariotos

superiores, coleções de mRNA nos quais as sequências intrônicas foram removidas por spliceossomos são um ponto de início mais útil do que o DNA genômico. A sequência do mRNA pode ser virtualmente "traduzida" na sequência de aminoácidos da proteína simplesmente por meio da leitura dos códons triplos.

O **DNA complementar (cDNA)** é uma versão do DNA de uma molécula de mRNA. Os pesquisadores utilizam o cDNA em vez do próprio mRNA, tendo em vista que os RNA são inerentemente menos estáveis que o DNA. Além disso, o RNA não pode ser manipulado pelas enzimas disponíveis para a clonagem do DNA, e não existem técnicas para a amplificação e a purificação de rotina de moléculas de RNA individuais. O cDNA é produzido a partir do mRNA com a utilização de uma enzima especial, denominada *transcriptase reversa*, originalmente isolada de retrovírus (ver Capítulo 15). Os retrovírus apresentam genomas de RNA que são copiados no DNA que se insere no cromossomo do hospedeiro. Você consegue pensar o motivo de ela ser denominada transcriptase *reversa*? Para produzir o cDNA, um pesquisador purifica o mRNA de um tecido que produz uma grande quantidade da proteína desejada. A insulina é produzida nas células β das ilhotas do pâncreas, de modo que utilizaríamos aquele órgão como a nossa fonte para o mRNA da insulina. Em seguida, o mRNA purificado é adicionado a um tubo de ensaio contendo transcriptase reversa, os quatro dNTP, e um *primer* curto de resíduos de dTTP polimerizados (denominado *primer* oligo-dT). O *primer* oligo-dT une-se à cauda poli(A) da molécula de mRNA que está sendo copiada. Com a utilização dessa molécula de mRNA como um molde, a transcriptase reversa catalisa a síntese de uma molécula de DNA unifilamentar com início a partir do *primer* oligo-dT. Quando alcança a extremidade do molde do RNA, a transcriptase reversa sintetiza uma alça em grampo. Quando o mRNA é removido (por meio de tratamento com uma solução básica), a alça em grampo pode atuar como um *primer* natural para que a DNA polimerase copie o cDNA em uma molécula de DNA bifilamentar (Figura 10.4). Assim como os fragmentos do DNA genômico ou produtos da PCR, esse cDNA bifilamentar pode ser inserido em moléculas de DNA recombinante para amplificação adicional, ou pode ser utilizado em qualquer outro procedimento com base no DNA, conforme descrito neste capítulo.

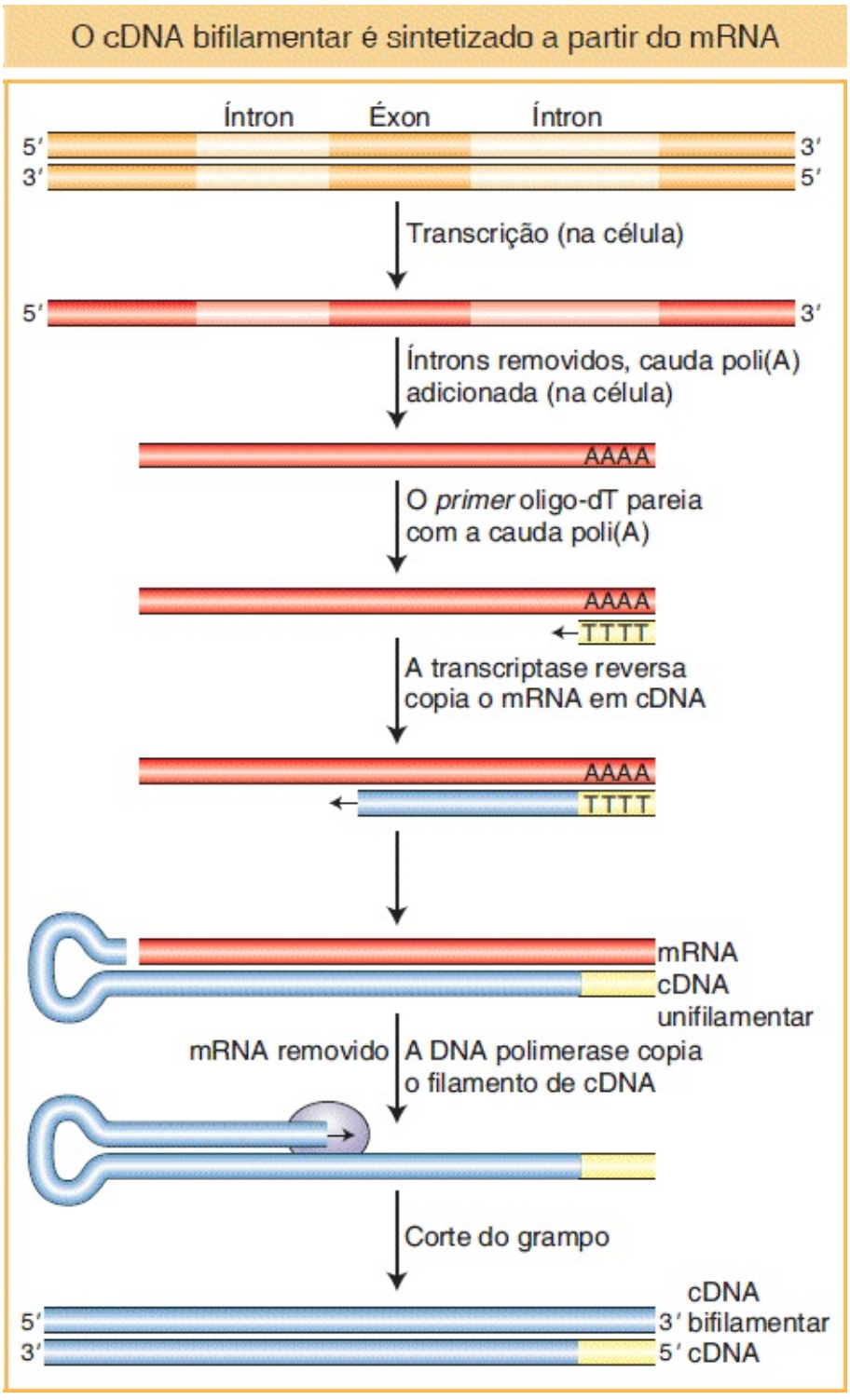


FIGURA 10.4 A formação do cDNA para o gene da insulina. O gene da insulina (com seus dois íntrons) é transcrito no pâncreas em pré-mRNA. O íntrons são removidos por *splicing* e resíduos A são adicionados à extremidade 3′ para formar o mRNA poliadenilado. Em laboratório, os mRNA são isolados de células pancreáticas e um *primer* oligo-dT curto é hibridizado com a cauda poli(a) de todos os mRNA para iniciar a

síntese de DNA complementar a partir do molde de RNA pela transcriptase reversa. A transcriptase reversa sintetiza uma estrutura de alça em grampo que atua como um *primer* para a síntese do segundo filamento de cDNA, após o filamento do mRNA ter sido degradado (por meio do tratamento com NaOH ou com RNAseH).

> CONCEITO-CHAVE O mRNA com frequência é um ponto de início preferível no isolamento de um gene. A conversão enzimática do mRNA em cDNA possibilita o isolamento de uma cópia de um gene sem íntrons.

Ligação do DNA doador e do vetor

Conforme descrito anteriormente, temos diversas opções para a obtenção do gene da insulina humana a partir do genoma, ou do mRNA purificado. Esses métodos produzem fragmentos de DNA genômico, produtos da PCR, ou cDNA bifilamentar. A próxima etapa é construir moléculas de DNA recombinante por meio da inserção do DNA doador no DNA vetor.

Clonagem de fragmentos de DNA com extremidades coesivas. Relembre que os cientistas originais que isolaram o gene da insulina humana não conheciam a sequência do gene e, assim, precisavam criar uma biblioteca de fragmentos de DNA do genoma humano a partir da qual isolariam o gene específico. Para produzir moléculas de DNA recombinante contendo fragmentos de DNA genômico doador, ambos os DNA, doador e vetor, são digeridos por uma enzima de restrição que produz as mesmas extremidades coesivas complementares (ver Figura 10.2). Os fragmentos resultantes em seguida são misturados para possibilitar que as extremidades coesivas do DNA doador e do vetor hibridizem entre si e formem moléculas recombinantes. A Figura 10.5 A demonstra um DNA de plasmídio bacteriano que carrega um sítio de restrição de *Eco*RI único, de modo que a digestão com a enzima de restrição *Eco*RI converte o DNA circular em uma molécula linear única com extremidades coesivas. O DNA doador de qualquer outra fonte, tal como o DNA humano, também é tratado com a enzima *Eco*RI para produzir uma população de fragmentos que carregam as mesmas extremidades coesivas. Quando as duas populações são misturadas sob condições

fisiológicas adequadas, os fragmentos de DNA das duas fontes podem hibridizar, tendo em vista que se formam duplas-hélices entre as suas extremidades coesivas (Figura 10.5 B). Em qualquer reação de clonagem, existem muitas moléculas de plasmídios linearizadas na solução, bem como muitos fragmentos diferentes de *Eco*RI do DNA doador, uma pequena fração dos quais apresentará o DNA-alvo. Portanto, será produzido um arranjo diverso de plasmídios recombinados com diferentes fragmentos doadores. Nesse estágio, as moléculas hibridizadas não apresentam arcabouços açúcar-fosfato unidos de modo covalente e provavelmente serão destruídas, tendo em vista que oito ligações de hidrogênio proporcionam apenas ligações fracas entre as sequências. Entretanto, os arcabouços podem ser covalentemente selados por meio da adição da enzima **DNA ligase**, que cria ligações fosfodiéster nas junções (Figura 10.5 C).

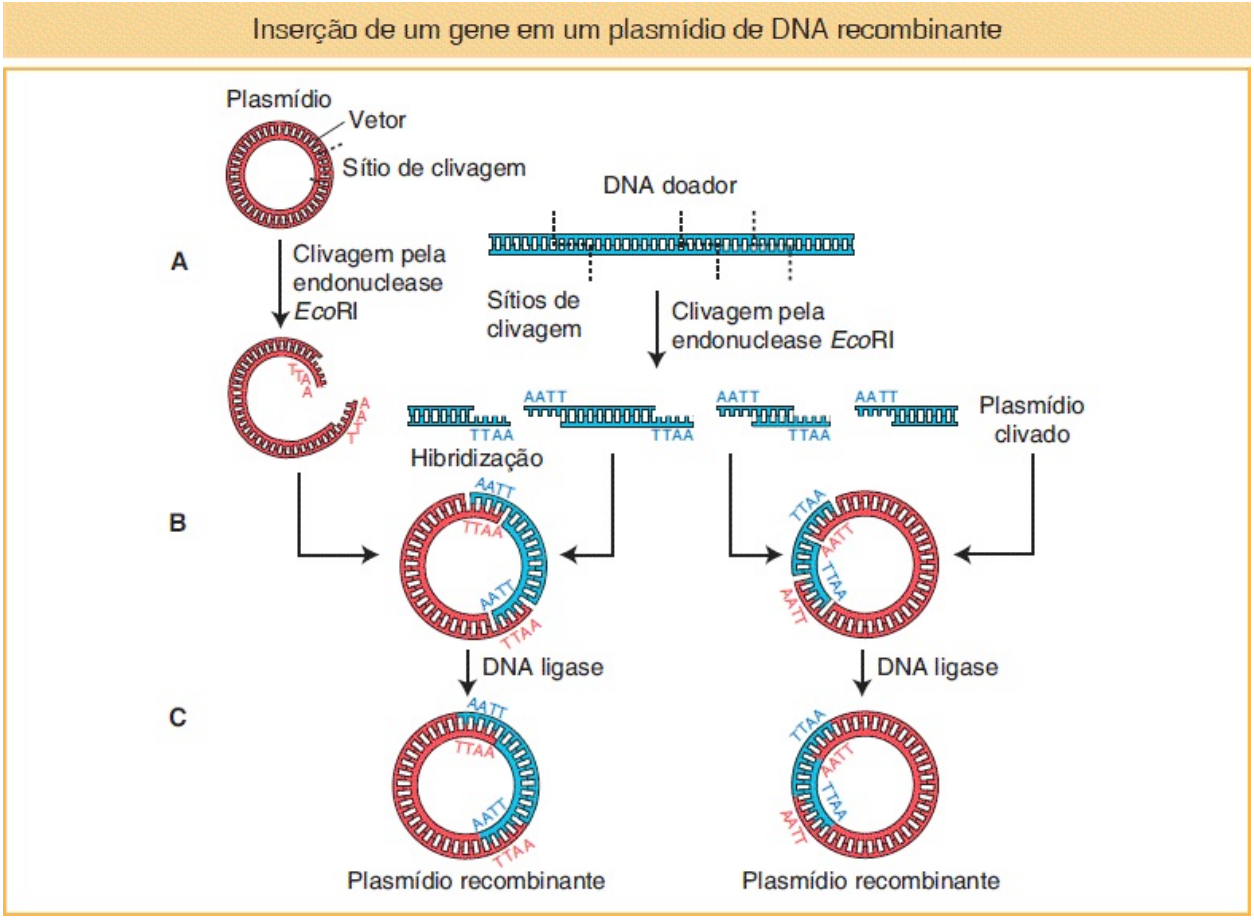


FIGURA 10.5 Método para a produção de uma coleção de plasmídios de DNA recombinante contendo genes derivados da digestão de DNA doador com enzima de restrição.

Clonagem de fragmentos de DNA com extremidades cegas. O conhecimento sobre a sequência do gene da insulina humana nos ajuda a focar no gene, mas adiciona uma pequena complicação na reação de clonagem. Algumas enzimas de restrição produzem extremidades cegas em vez de cortes alternados. Além disso, o cDNA e os fragmentos de DNA que têm origem a partir da PCR apresentam extremidades cegas ou quase cegas. Enquanto os fragmentos com extremidades cegas de todas essas fontes podem ser unidos ao vetor apenas com a utilização da ligase, essa é uma reação muito ineficiente, tendo em vista que as extremidades cegas não podem ficar juntas. Um método alternativo é criar produtos de PCR com extremidades coesivas por meio da utilização de *primers* de PCR especialmente desenhados que contêm sequências de reconhecimento de endonuclease de restrição nas suas extremidades 5′ (Figura 10.6). A digestão do produto final da PCR com a enzima de restrição (*Eco*RI nesse caso) produz um fragmento que está pronto para ser inserido em um vetor (ver Figura 10.5 B).

Outro método adiciona extremidades coesivas a qualquer fragmento de DNA bifilamentar — incluindo os cDNA (Figura 10.7). Oligonucleotídios bifilamentares curtos (denominados ligadores ou adaptadores) que contêm um sítio de restrição são adicionados a um tubo de ensaio que contém cDNA e ligase. A ligase une os ligadores às extremidades dos filamentos de cDNA. Após a ligação estar completa, o DNA é incubado com a enzima de restrição correspondente para gerar as extremidades coesivas necessárias para a clonagem em um vetor plasmidial (ver Figura 10.5 B). Observe que nos exemplos demonstrados, ambos o DNA amplificado e o cDNA não devem conter um sítio de *Eco*RI interno, ou ele também será digerido. Se ele contiver, a sequência para um sítio de restrição que não está no DNA amplificado pode ser adicionada aos *primers* ou aos ligadores.

> CONCEITO-CHAVE DNA doadores e de vetores com as mesmas extremidades coesivas podem ser unidos e ligados com eficiência. Alternativamente, o DNA doador que é o produto da PCR ou da síntese de cDNA requer a adição de extremidades coesivas antes da inserção em um

> vetor.

Amplificação do DNA doador em uma célula bacteriana

A amplificação das moléculas de DNA recombinantes tira vantagem de processos genéticos procarióticos tais como a transformação bacteriana, a replicação de plasmídios e o crescimento de bacteriófagos, todos discutidos no Capítulo 5. A Figura 10.8 ilustra a clonagem de um segmento de DNA doador. Um vetor recombinante único entra em uma célula bacteriana e é amplificado pelo mesmo maquinário que replica o cromossomo bacteriano. Uma exigência básica é a presença de uma origem de replicação do DNA reconhecida pelas proteínas de replicação do hospedeiro (conforme descrito no Capítulo 7). Logo existem muitas cópias de cada vetor em cada célula bacteriana. Portanto, após a amplificação, uma colônia de bactérias conterá tipicamente bilhões de cópias do inserto de DNA doador único fundido ao seu vetor. Esse conjunto de cópias amplificadas do fragmento de DNA doador único dentro do vetor de clonagem é o *clone* de DNA recombinante.

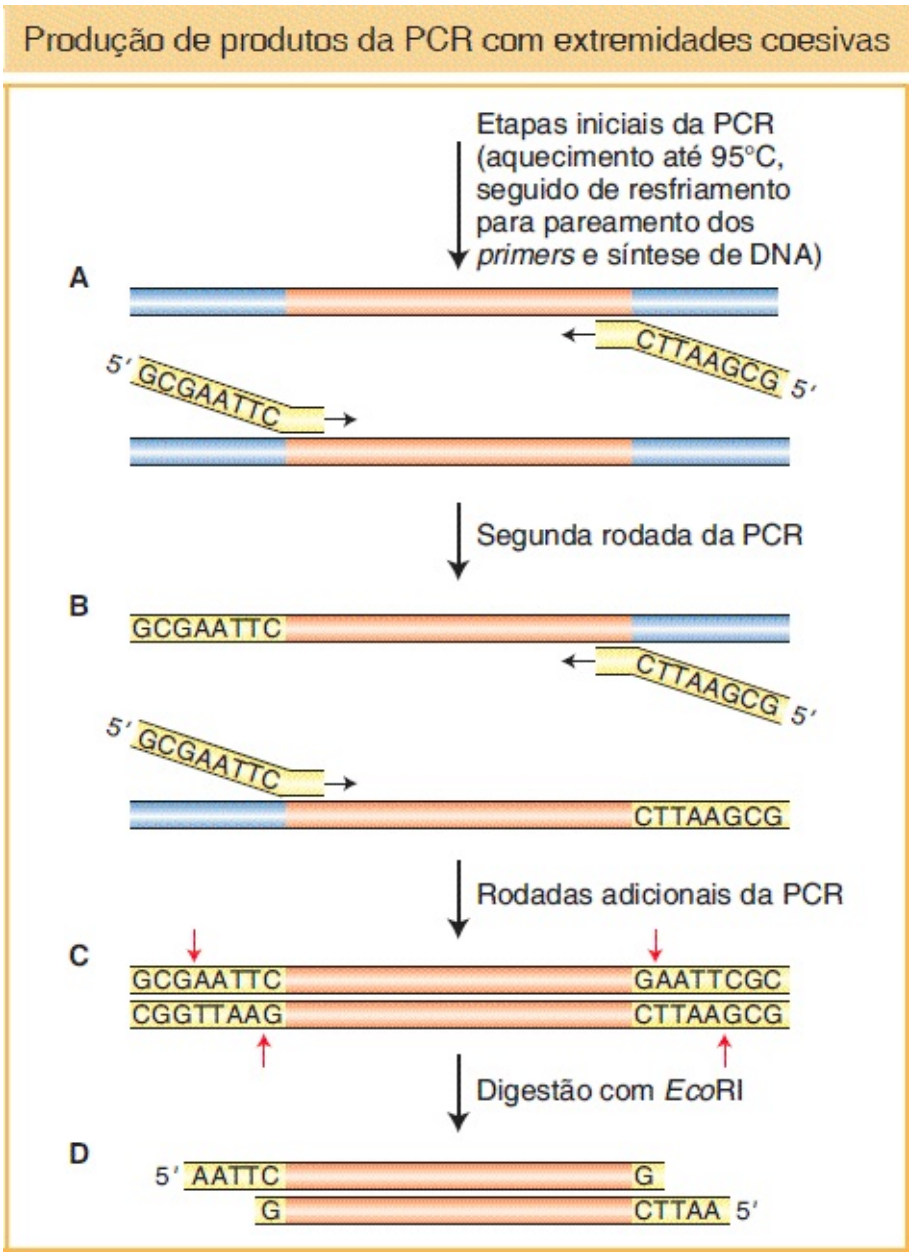


FIGURA 10.6 Adição de sítios de *Eco*RI nas extremidades de produtos da PCR. **A.** Um par de *primers* de PCR é desenhado de modo que as suas extremidades 3′ unam-se à sequência-alvo, enquanto suas extremidades 5′ contenham sequências que codificam o sítio da enzima de restrição (*Eco*RI neste caso). Dois nucleotídios adicionais (aleatórios) são adicionados à extremidade 5’, tendo em vista que as enzimas de restrição necessitam de sequências em ambos os lados da sequência de reconhecimento para o corte eficiente. O DNA-alvo é desnaturado, e as extremidades 5′ com os sítios de restrição permanecem unifilamentares, enquanto o restante dos *primers* pareia e é estendido pela DNA polimerase. **B.** Na segunda rodada da PCR — apenas os filamentos recém-sintetizados estão demonstrados — os *primers* de DNA pareiam novamente, e dessa vez a síntese do DNA produz moléculas de DNA bifilamentares como a PCR convencional, mas essas moléculas apresentam sítios de restrição em uma extremidade. **C.** Os produtos da segunda rodada e de todas

as rodadas subsequentes apresentam sítios de *Eco*RI em ambas as extremidades. **D.** Quando estas são cortadas com a *Eco*RI, são produzidas extremidades coesivas.

A amplificação do DNA doador dentro de uma célula bacteriana envolve as etapas a seguir:

- Escolha de um vetor de clonagem e introdução do inserto (ver seção antecedente para uma discussão sobre a última)
- Introdução da molécula de DNA recombinante em uma célula bacteriana
- Recuperação das moléculas recombinantes amplificadas.

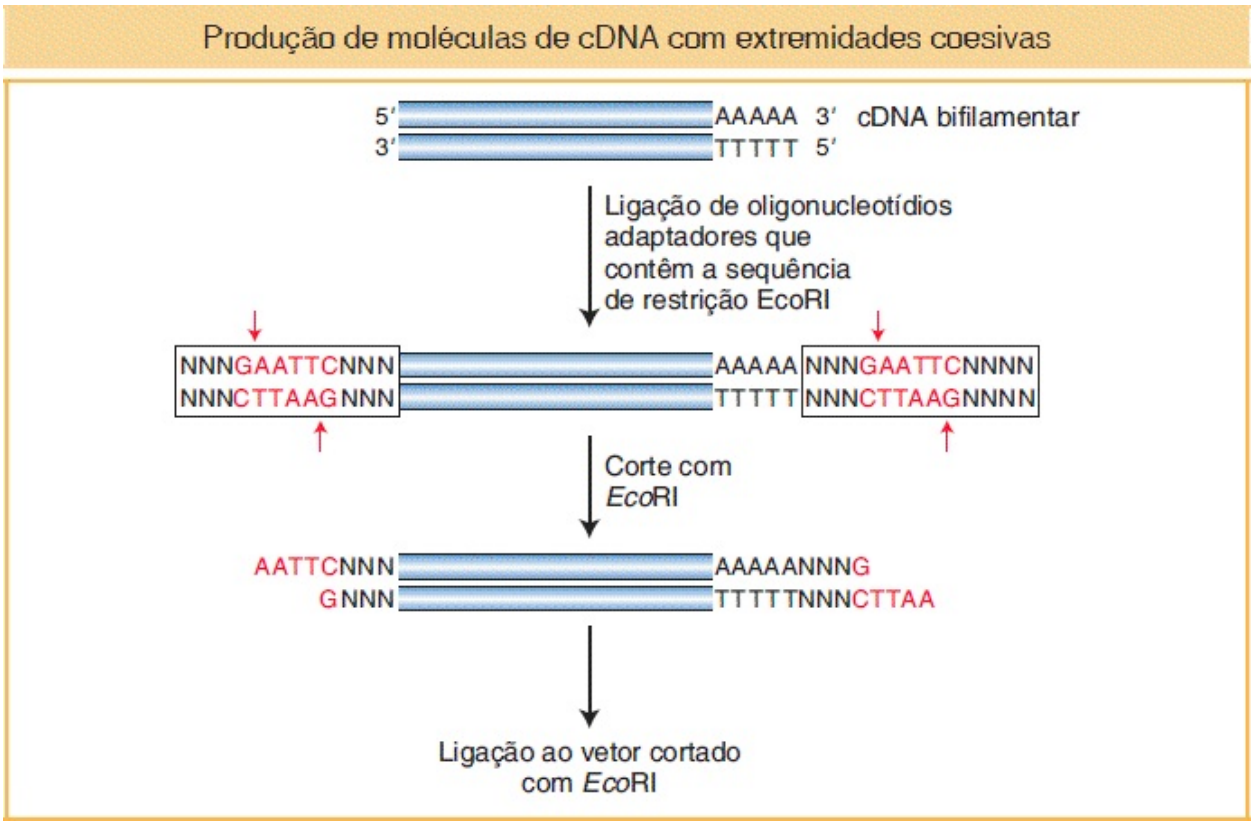


FIGURA 10.7 Adição de sítios de *Eco*RI às extremidades de moléculas de cDNA. As moléculas de cDNA têm origem na última etapa na Figura 10.4. Adaptadores (*regiões nos quadros*) são adicionadas a ambas as extremidades das moléculas de cDNA. Esses adaptadores são oligonucleotídios bifilamentares que contêm um sítio de restrição (*o sítio* Eco*RI está demonstrado em vermelho*) e uma sequência de DNA aleatória em ambas as extremidades (*representada por N*).

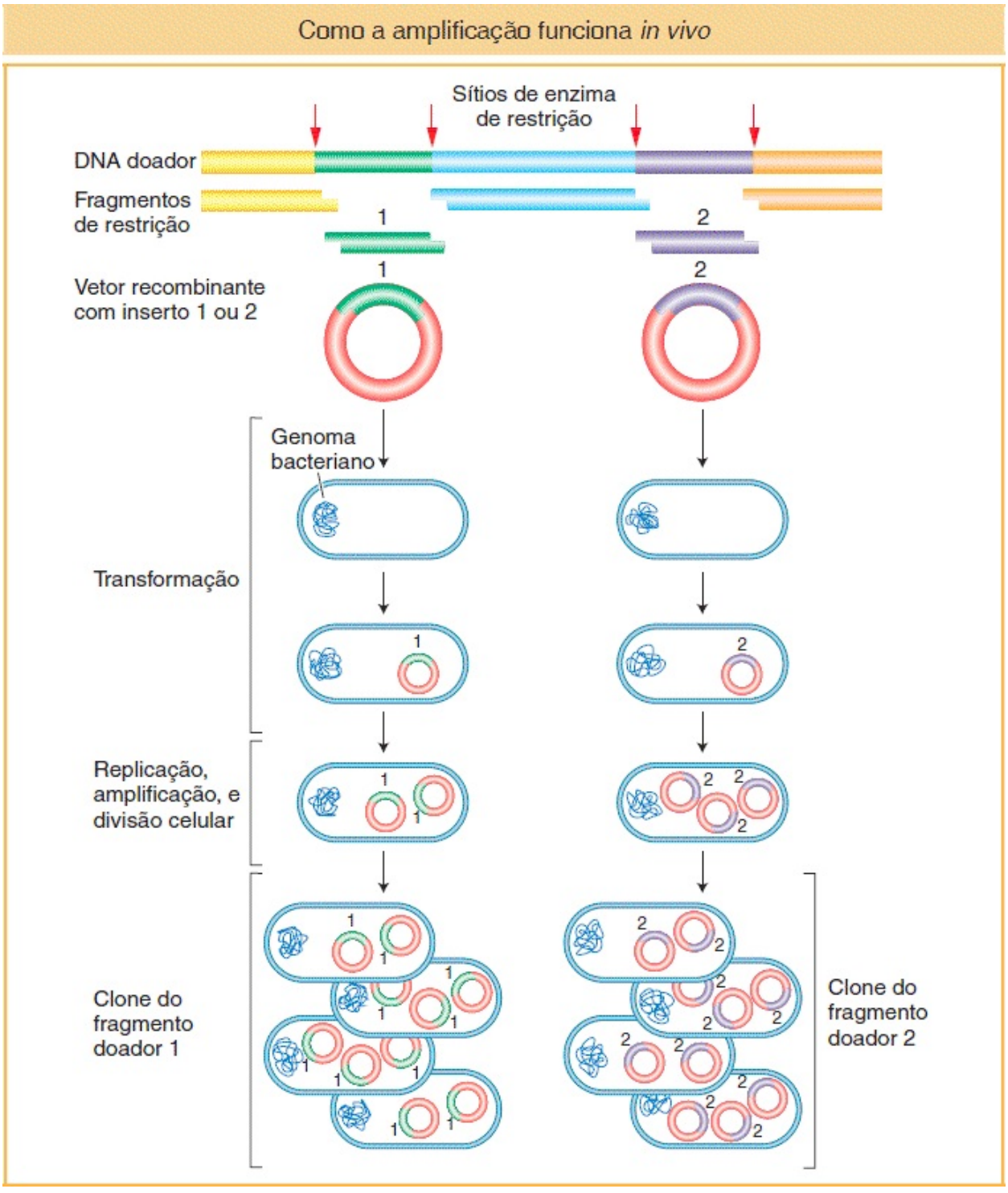


FIGURA 10.8 A estratégia geral utilizada para clonar um gene. O tratamento com enzima de restrição do DNA doador e do vetor possibilita a inserção de fragmentos únicos nos vetores. Um vetor único entra em um hospedeiro bacteriano, no qual a replicação e a divisão celulares resultam em um grande número de cópias do fragmento doador.

Escolha dos vetores de clonagem. Os vetores devem ser moléculas pequenas para manipulação conveniente, mas podem variar de diversas formas para a adequação ao objetivo do experimento. Alguns vetores devem ser capazes de replicação prolífica em uma célula viva com a finalidade de amplificar o fragmento doador inserido. Contrariamente, outros são projetados para estarem presentes em apenas uma única cópia para manter a integridade do DNA inserido (ver adiante). Todos os vetores devem apresentar sítios de restrição convenientes, nos quais o DNA a ser clonado possa ser inserido (denominado um sítio poliligador ou de clonagem múltipla). Idealmente, o sítio de restrição deve estar presente apenas uma vez no vetor, tendo em vista que os fragmentos de restrição do DNA doador serão inseridos apenas naquela localização única no vetor. Também é importante ter um modo de identificar e recuperar rapidamente a molécula recombinante desejada. Diversos vetores de clonagem que atendem uma ampla diversidade de necessidades experimentais estão em uso atualmente. Em seguida estão algumas classes gerais de vetores de clonagem.

Vetores plasmidiais. Conforme descrito anteriormente, os plasmídios bacterianos são pequenas moléculas circulares de DNA que comumente replicam seu DNA independentemente do cromossomo bacteriano. Os plasmídios rotineiramente utilizados como vetores carregam um gene de resistência a fármacos e um gene para distinguir plasmídios com e sem insertos de DNA. Esses genes de resistência a fármacos proporcionam um modo conveniente de selecionar células bacterianas transformadas por plasmídios: aquelas células ainda vivas após a exposição ao fármaco devem carrear os vetores plasmidiais. Entretanto, nem todos os plasmídios nessas células transformadas conterão insertos de DNA. Alguns vetores plasmidiais também apresentam um sistema que possibilita que os pesquisadores identifiquem colônias bacterianas com plasmídios que contenham insertos de DNA. Por esse motivo, é desejável ser capaz de identificar colônias bacterianas com plasmídios que contenham insertos de DNA. Uma referida característica é parte do vetor plasmidial pUC18 demonstrado na Figura 10.9; os insertos de DNA rompem um gene (*lacZ*) no plasmídio que codifica uma enzima (β-galactosidase) necessária para clivar um composto adicionado ao ágar da

placa de Petri (X-gal), de modo que ele produza um pigmento azul. Portanto, as colônias que contêm os plasmídios com o inserto de DNA serão brancas, em vez de azuis (elas não conseguem clivar o X-gal, tendo em vista que não produzem β-galactosidase).

Vetores bacteriófagos. Um vetor bacteriófago abriga DNA como um inserto "compactado" dentro da partícula do fago. Diferentes classes de vetores bacteriófagos podem carregar insertos de DNA doador de diferentes tamanhos. O bacteriófago λ (lambda; discutido nos Capítulo 5 e 11) é um vetor de clonagem eficaz para insertos de DNA bifilamentares tão longos quanto 15 kb. A parte central do genoma do fago não é necessária para a replicação ou o acondicionamento de moléculas de DNA λ em *E. coli* e, assim, pode ser cortada por enzimas de restrição e descartada. A parte central deletada em seguida é substituída por insertos de DNA doador.

Utilização de um vetor plasmidial, pUC18

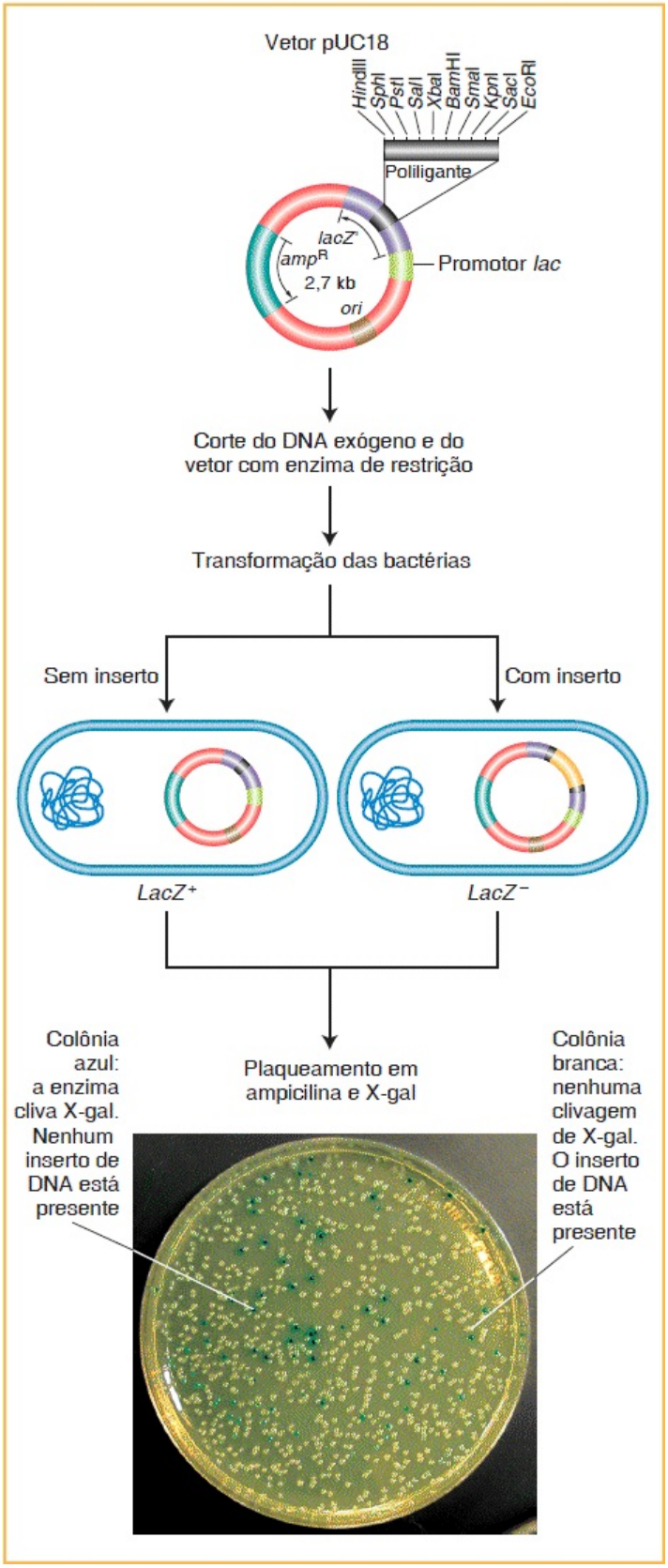


FIGURA 10.9 O vetor plasmidial pUC18 foi designado para a utilização como um vetor para a clonagem de DNA. A inserção de DNA em pUC18 é detectada por meio da inativação da função da β-galactosidase de *lacZ*, resultando em uma capacidade de converter o substrato artificial X-gal em um corante azul. O poliligador apresenta diversos sítios de restrição alternativos, dentro dos quais o DNA doador pode ser inserido. (Foto: Dr. James M. Burnette III e Dra. Leslie Bañuelos.)

Vetores para insertos de DNA maiores. Os vetores plasmidiais padrão e o fago λ já descritos podem aceitar DNA doador de tamanhos tão grandes quanto 10 a 15 kb. Entretanto, muitos experimentos necessitam de insertos que excedem em muito esse limite superior. Para atender a essas necessidades, foram construídos por engenharia genética vetores especiais que necessitam de métodos mais sofisticados para a transferência de DNA para a célula hospedeira. Em todos os casos, os DNA replicam como plasmídios grandes após terem sido introduzidos na bactéria.

Fosmídios são vetores que conseguem carrear insertos de 35 a 45 kb (Figura 10.10). Eles são híbridos modificados de DNA de fago λ e DNA de plasmídio F bacteriano (ver Capítulo 5). Os fosmídios são acondicionados nas partículas de fago λ, que atuam como as seringas que introduzem esses grandes pedaços de DNA recombinante dentro de células de *E. coli* receptoras. Após estarem na célula, esses híbridos, assim como o fago λ, formam moléculas circulares que replicam extracromossomicamente de modo semelhante aos plasmídios. Entretanto, em virtude da presença de origens de replicação do plasmídio F que acoplam a replicação do plasmídio à duplicação do cromossomo da célula hospedeira, muito poucas cópias de fosmídios se acumulam em uma célula.

O vetor mais popular para a clonagem de insertos de DNA muito grandes em bactérias é o **cromossomo artificial bacteriano (BAC)**. Derivado do plasmídio F, ele pode carregar insertos que variam de 100 a 200 kb, embora o próprio vetor seja de apenas aproximadamente 7 kb (ver Figura 10.10). O DNA a ser clonado é inserido no plasmídio, e esse grande DNA recombinante circular é introduzido na bactéria. Os BAC foram os vetores "burros de carga" para a extensiva clonagem requerida por projetos de sequenciamento genômico em grande escala, incluindo o projeto público para sequenciar o genoma humano (discutido no Capítulo 14).

> CONCEITO-CHAVE O conjunto de ferramentas do engenheiro genético contém uma diversidade de vetores de clonagem que aceitam insertos de tamanhos pequenos em plasmídios, de tamanhos médios em bacteriófagos, e de tamanhos grandes em fosmídios e BAC.

Entrada das moléculas recombinantes na célula bacteriana. São utilizados três métodos para introduzir moléculas de DNA recombinante nas células bacterianas: transformação, transdução e infecção (Figura 10.11; ver seções 5.3 e 5.4).

- Na transformação, as bactérias são banhadas em uma solução que contém a molécula de DNA recombinante. Tendo em vista que as células bacterianas utilizadas em pesquisas não conseguem captar moléculas de DNA tão grandes quanto os plasmídios recombinantes, elas devem se tornar *competentes* (ou seja, capazes de captar o DNA do meio circundante) por meio da incubação em uma solução de cálcio ou da exposição a um pulso elétrico de alta voltagem (*eletroporação*). Após a entrada em uma célula competente através dos poros da membrana, a molécula recombinante se torna um cromossomo de plasmídio (Figura 10.11 A). A eletroporação é o método de escolha para a introdução de DNA especialmente grandes, tais como BAC, nas células bacterianas

Fosmídios e BAC são vetores de clonagem que carregam insertos grandes

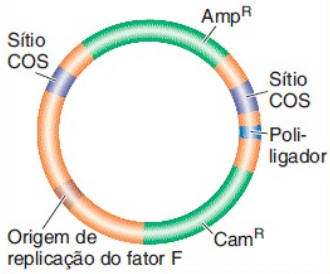
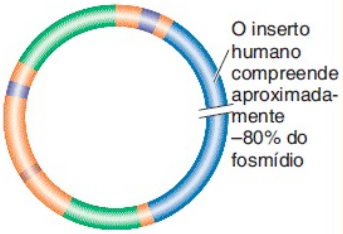
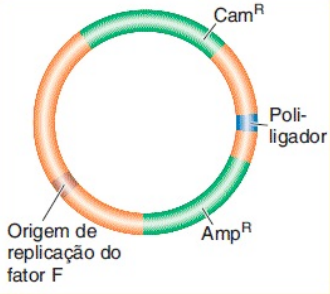
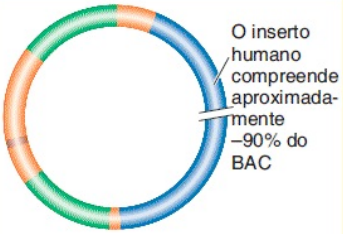
	Vetor	Tamanho do DNA genômico parcialmente digerido	Clone de inserto grande	Número de cópias	Número de clones para abranger 1× o genoma humano
Fosmídio		35 a 45 kb		Cópia única	Aproximadamente 75.000
BAC		100 a 200 kb		Cópia única	15.000 a 30.000

FIGURA 10.10 Características de alguns vetores de clonagem de inserto grande. O número de clones necessário para abranger o genoma humano uma vez (1×) tem por base um tamanho de genoma de 3.000 Mb (3 bilhões de pares de bases).

Modos de administração de DNA recombinante dentro das células bacterianas

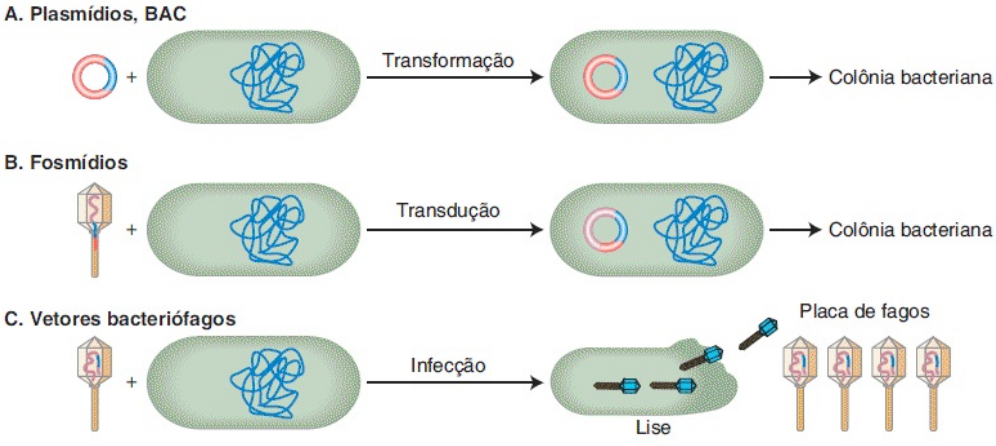


FIGURA 10.11 O DNA recombinante pode ser inserido nas células bacterianas por meio de transformação, transdução, ou infecção com um fago. **A.** Vetores plasmidiais e BAC são inseridos por transformação mediada por DNA. **B.** Determinados vetores, tais como os fosmídios, são inseridos nas cabeças dos bacteriófagos (transdução); entretanto, após terem sido injetados na bactéria, eles formam círculos e são replicados como grandes plasmídios. **C.** Vetores bacteriófagos, tais como o fago λ, infectam e lisam a bactéria, liberando um

clone de fagos como progênie, todos carregando a molécula de DNA recombinante idêntica no genoma do fago.

- Na transdução, a molécula recombinante é combinada com a cabeça do fago e as proteínas da cauda para produzir um vírus que contém um DNA amplamente não viral. Esses fagos modificados em seguida são misturados com bactérias e injetam sua carga de DNA nas células bacterianas, mas novos fagos não conseguem se formar, tendo em vista que eles não carregam os genes virais necessários para a replicação dos fagos. Os fosmídios são introduzidos dentro das células por meio de transdução (Figura 10.11 B)
- Contrariamente à transdução, que produz plasmídios e colônias bacterianas, mas não novos vírus, a infecção produz partículas de fagos recombinantes (Figura 10.11 C). Por meio de rodadas repetidas de reinfecção, uma placa repleta de partículas de fagos é formada a partir de cada bactéria inicial que foi infectada. Cada partícula de fago em uma placa contém não apenas o DNA recombinante, mas também os genes virais necessários para criar novas partículas de fagos infecciosas.

Recuperação das moléculas recombinantes amplificadas. O DNA recombinante empacotado dentro das partículas de fagos é facilmente obtido por meio da coleta do lisado de fago e do isolamento do DNA que ele contém. Para obter o DNA recombinante acondicionado em plasmídios, fosmídios, ou BAC, as bactérias são quebradas química ou mecanicamente. O plasmídio de DNA recombinante é separado do cromossomo bacteriano principal muito maior por meio de centrifugação, eletroforese ou outras técnicas seletivas que distinguem o cromossomo do plasmídio pelo tamanho ou pela forma.

> CONCEITO-CHAVE A clonagem de genes é realizada por meio da introdução de vetores recombinantes únicos em células bacterianas receptoras, seguida pela amplificação dessas moléculas como cromossomos de plasmídios ou fagos.

Construção de bibliotecas genômicas e de cDNA

Vimos como obter e amplificar moléculas de DNA recombinante individuais, tais como o nosso cDNA de insulina humana. Considere a tarefa em 1982, quando o gene da insulina humana deveria ser identificado a partir de uma biblioteca de fragmentos do genoma humano. Para assegurar que clonamos o segmento de DNA de interesse, precisamos produzir grandes coleções de segmentos de DNA que sejam totalmente inclusivas. Por exemplo, tomamos todo o DNA de um genoma, quebramos em segmentos do tamanho correto para o nosso vetor de clonagem e inserimos cada segmento em uma cópia diferente do vetor, criando, assim, uma coleção de moléculas de DNA recombinante que, consideradas em conjunto, representam o genoma inteiro. Em seguida, transformamos ou infectamos essas moléculas em células receptoras bacterianas em separado, onde elas são amplificadas. A coleção resultante de bactérias ou bacteriófagos que contêm o DNA recombinante é denominada **biblioteca genômica**. Se estivermos utilizando um vetor de clonagem que aceita um inserto de tamanho médio de 10 kb e se o genoma inteiro for do tamanho de 100.000 kb (o tamanho aproximado do genoma do nematódeo *Caenorhabditis elegans*), então no mínimo 10.000 clones recombinantes independentes serão necessários para representar o equivalente a um genoma de DNA. Para assegurar que todas as sequências genômicas que podem ser clonadas estejam contidas em uma coleção, as bibliotecas genômicas tipicamente representam um segmento médio do genoma no mínimo cinco vezes (e assim, em nosso exemplo, haverá 50.000 clones independentes na biblioteca genômica). Essa representação múltipla torna altamente improvável que, ao acaso, uma sequência não esteja representada no mínimo uma vez na biblioteca.

De modo semelhante, coleções representativas de insertos de cDNA requerem dezenas ou centenas de milhares de clones de cDNA independentes; essas coleções são **bibliotecas de cDNA** e representam apenas as regiões codificadoras de proteínas do genoma. Uma biblioteca de cDNA abrangente inclui amostras de mRNA de diferentes tecidos, diferentes estágios do desenvolvimento ou de organismos cultivados em diferentes condições ambientais.

Se optamos por construir uma biblioteca de DNA genômico ou uma biblioteca

de cDNA depende da situação. Se estivermos procurando um gene específico que seja ativo em um tipo específico de tecido em uma planta ou um animal, então faz sentido construir uma biblioteca de cDNA a partir de uma amostra daquele tecido. Por exemplo, suponha que desejamos identificar cDNA que correspondam aos mRNA da insulina. As células β das ilhotas do pâncreas são a fonte de insulina mais abundante e, assim, os mRNA das células pancreáticas são a fonte apropriada para uma biblioteca de cDNA, tendo em vista que esses mRNA devem ser enriquecidos para o gene em questão. Uma biblioteca de cDNA representa um subconjunto das regiões transcritas do genoma; assim, ela inevitavelmente será menor do que uma biblioteca genômica completa. Embora as bibliotecas genômicas sejam maiores, elas apresentam o benefício de conter genes em sua forma nativa, incluindo íntrons e sequências reguladoras não transcritas. Uma biblioteca genômica é necessária em algum estágio como um prelúdio para a clonagem de um gene inteiro ou de um genoma inteiro.

> CONCEITO-CHAVE A tarefa de isolamento de um clone de um gene específico tem início com a produção de uma biblioteca de DNA genômico ou cDNA — se possível, enriquecida em relação às sequências que contêm o gene em questão.

10.3 Utilização de sondas moleculares para encontrar e analisar um clone de interesse específico

A produção de uma biblioteca conforme acabamos de descrever por vezes é denominada clonagem *shotgun*, tendo em vista que o experimentador clona uma grande amostra de fragmentos e espera que um dos clones contenha um “acerto” — o gene desejado. A tarefa em seguida é encontrar aquele clone em particular, considerada em seguida.

Detecção de clones específicos com a utilização de sondas

Uma biblioteca pode conter até centenas de milhares de fragmentos de DNA clonados. Essa enorme coleção de fragmentos deve ser triada para encontrar a molécula de DNA recombinante que contém o gene de interesse para um pesquisador. A referida triagem é realizada por meio da utilização de uma **sonda** específica, que encontrará e marcará apenas o clone desejado. Existem dois tipos de sondas: (1) aquelas que reconhecem uma sequência de ácido nucleico específica e (2) aquelas que reconhecem uma proteína específica.

Sondas para detectar DNA. As sondas para DNA fazem uso do poder da complementaridade de bases. Dois ácidos nucleicos unifilamentares com sequência de bases complementares total ou parcial "encontrarão" um ao outro em uma solução por colisão aleatória. Após a sua união, o híbrido bifilamentar assim formado é estável. Essa abordagem proporciona um meio poderoso para a detecção de sequências de interesse específicas. O uso de sondas de DNA requer que todas as moléculas se tornem unifilamentares por meio de aquecimento. Uma sonda unifilamentar marcada radioativa ou quimicamente é enviada para detectar sua sequência-alvo complementar em uma população de DNA tal como uma biblioteca. Sondas tão pequenas quanto de 15 a 20 pares de bases hibridizarão com as sequências complementares específicas dentro de DNA clonados muito maiores.

A identificação de um clone específico em uma biblioteca é um procedimento de múltiplas etapas. Na Figura 10.12, essas etapas estão demonstradas em relação a uma biblioteca clonada em um vetor fosmídio. As etapas são semelhantes para bibliotecas de plasmídios ou BAC. Em relação às bibliotecas de fagos, placas são triadas em vez de colônias. Primeiramente, colônias da biblioteca em uma placa de Petri são transferidas para uma membrana absorvente simplesmente posicionando-se a membrana sobre a superfície do meio. A membrana é retirada, as colônias aderidas à superfície da membrana são lisadas no local e o DNA é simultaneamente desnaturado, de modo a se tornar unifilamentar. Em segundo lugar, a membrana é banhada com uma solução de uma sonda unifilamentar que é específica para a sequência de DNA que está sendo procurada. Em geral, a própria sonda é um fragmento de DNA clonado, que apresenta uma sequência que

é complementar àquela do gene desejado. A sonda deve ser marcada com um isótopo radioativo ou um corante fluorescente. Portanto, a posição de uma marcação radioativa ou fluorescente concentrada indicará a posição do clone positivo. Em relação às sondas radioativas, a membrana é posicionada sobre um pedaço de filme de raios X e o decaimento do radioisótopo produz partículas subatômicas que “expõem” o filme, produzindo nele um ponto escuro adjacente à localização da concentração do radioisótopo. Um referido filme exposto é denominado **autorradiograma**. Se um corante fluorescente for utilizado como um marcador, a membrana é exposta ao comprimento de onda de luz correto para ativar a fluorescência do corante, e é realizada uma fotografia da membrana para registrar a localização do corante fluorescente.

Detecção do clone de interesse por meio da utilização de sondas de DNA ou RNA

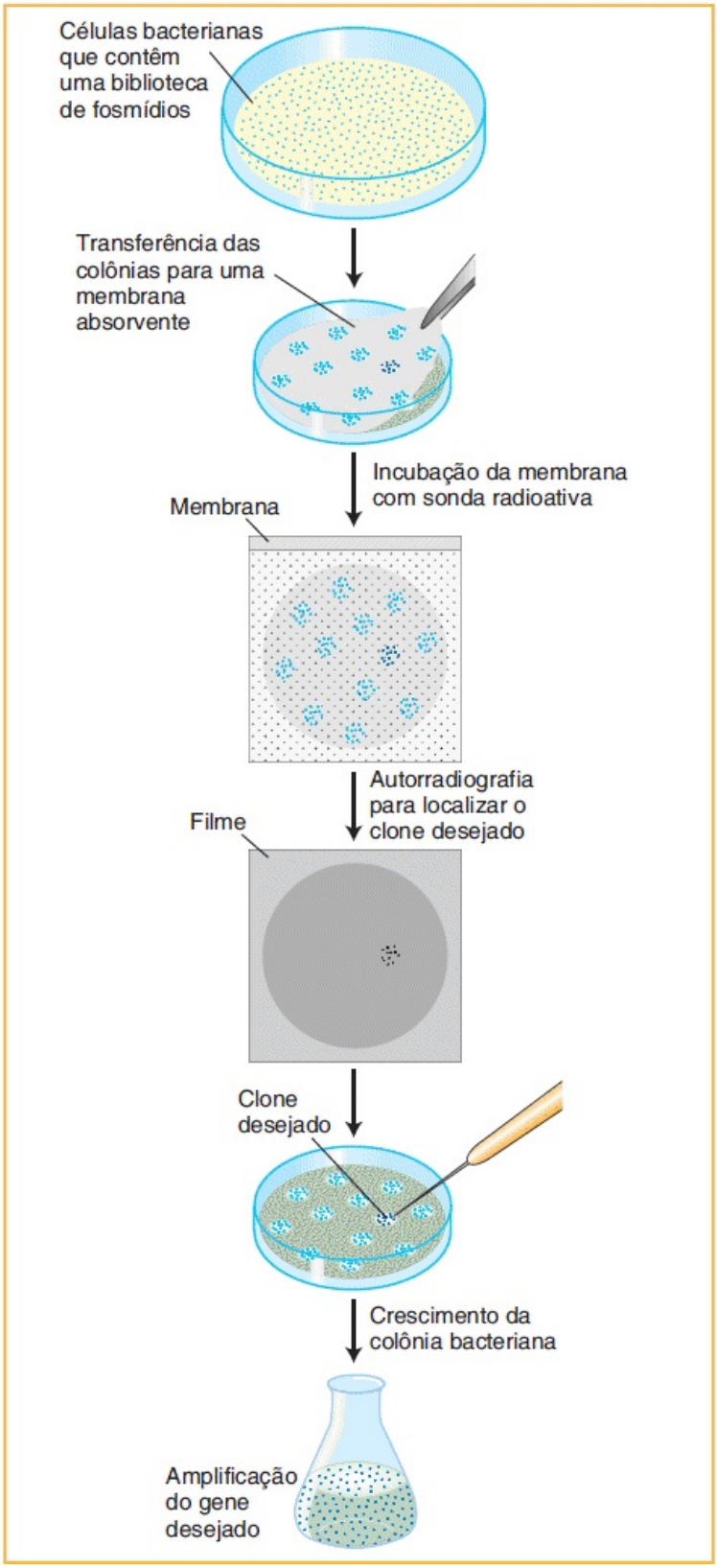


FIGURA 10.12 O clone que carrega um gene de interesse é identificado por meio da triagem de uma biblioteca genômica, neste caso produzida por meio da clonagem de genes em um vetor fosmídio, com DNA ou RNA sabidamente relacionado com o gene desejado. Uma sonda radioativa hibridiza com qualquer DNA recombinante que incorpore uma sequência de DNA complementar, e a posição do clone que apresenta o DNA é revelada por meio de autorradiografia. Agora o clone desejado pode ser selecionado a partir do ponto correspondente na placa de Petri e transferido para um novo hospedeiro bacteriano, de modo que um gene puro possa ser produzido.

Qual a origem do DNA para produzir uma sonda? O DNA pode ter origem em uma de diversas fontes.

- Pode-se utilizar um gene homólogo ou um cDNA de um organismo correlato. Esse método depende do fato de que organismos descendentes de um ancestral comum recente apresentarão sequências de DNA semelhantes. Embora o DNA da sonda e o DNA do clone desejado possam não ser idênticos, com frequência eles são semelhantes o suficiente para promover a hibridização.
- Pode-se utilizar o produto proteico do gene de interesse. Se toda a sequência da proteína, ou parte dela, for conhecida, pode-se retrotraduzi-la, por meio da utilização da tabela do código genético ao contrário (de aminoácido para códon), para deduzir as sequências de DNA que podem tê-la codificado. Em seguida, é projetada uma sonda de DNA sintética que corresponde àquela sequência. Relembre, entretanto, que o código genético é degenerado — ou seja, a maior parte dos aminoácidos é codificada por múltiplos códons. Portanto, teoricamente, diversas possíveis sequências de DNA podem codificar a proteína em questão, mas apenas uma dessas sequências de DNA está presente no gene que de fato codifica a proteína. Para contornar esse problema, é selecionado um trecho curto de aminoácidos com redundância mínima. Um conjunto misto de sondas é projetado em seguida, contendo todas as possíveis sequências de DNA que podem codificar essa sequência de aminoácidos. Esse "coquetel" de oligonucleotídios é utilizado como uma sonda, e um filamento correto ou muito semelhante nesse coquetel hibridizará com o gene de interesse. Oligonucleotídios de aproximadamente 20 nucleotídios de comprimento têm especificidade suficiente para hibridizar com uma única sequência de DNA

complementar na biblioteca.

Sondas para a detecção de proteínas. Se o produto proteico de um gene for conhecido e isolado na forma pura, essa proteína então poderá ser utilizada para detectar o clone do gene correspondente em uma biblioteca. O processo, descrito na Figura 10.13, requer dois componentes. Primeiramente, ele requer uma biblioteca de expressão, produzida por meio da utilização de vetores de expressão que direcionarão a célula hospedeira para produzir a proteína. Para construir a biblioteca, o cDNA é inserido no vetor na trinca correta do quadro de leitura com um promotor bacteriano e as células que contêm o vetor e seu inserto produzem uma tradução do inserto de cDNA. Em segundo lugar, o processo necessita de um anticorpo que se ligue ao produto proteico específico do gene de interesse. (Um **anticorpo** é uma proteína produzida pelo sistema imune de um animal que se liga com alta afinidade a uma determinada molécula.) O anticorpo é utilizado para triar a biblioteca de expressão daquela proteína. Uma membrana é sobreposta na superfície do meio e removida de modo que algumas das células de cada colônia agora estejam aderidas à membrana em locais que correspondem às suas posições sobre a placa de Petri original (ver Figura 10.13). A membrana impressa em seguida é seca e banhada em uma solução do anticorpo, que se ligará à impressão de qualquer colônia que contenha a proteína de fusão de interesse. Clones positivos são revelados por meio de um anticorpo secundário marcado que se liga ao primeiro anticorpo. Ao detectar a proteína correta, o anticorpo identifica o clone que contém o gene que deve ter sintetizado aquela proteína e que, portanto, contém o cDNA desejado.

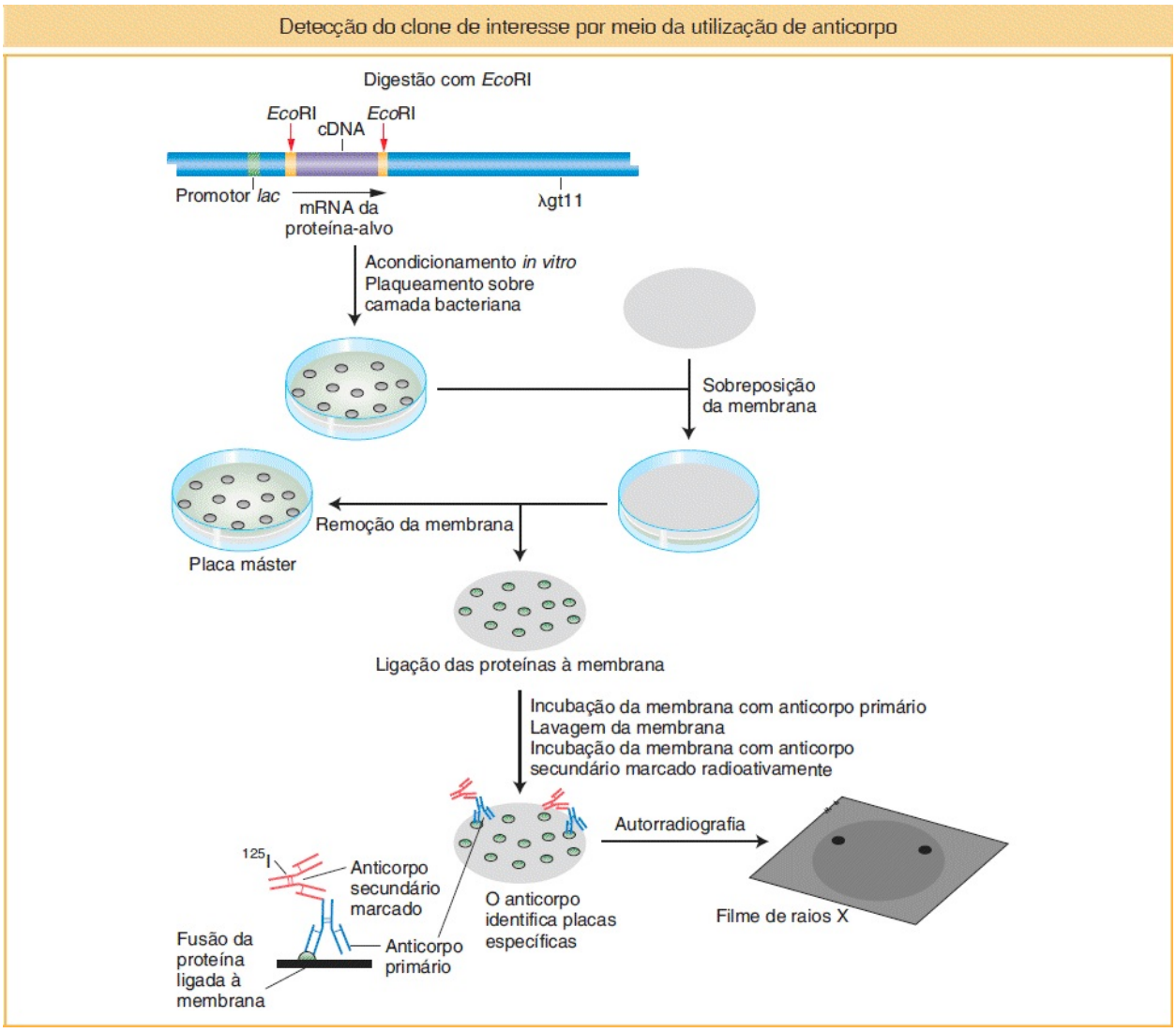


FIGURA 10.13 Para encontrar o clone de interesse, uma biblioteca de expressão construída com o vetor de fago λ especial denominado λgt11 é triada com um anticorpo específico de proteína. Após os anticorpos não ligados à membrana terem sido removidos, os anticorpos ligados são visualizados por meio da ligação de um anticorpo secundário radioativo.

Podemos observar como esse tipo de sonda atua na prática ao retornar ao exemplo da insulina humana. Para clonar um cDNA correspondente à insulina humana, primeiramente sintetizamos cDNA com a utilização de mRNA isolado de células pancreáticas como o molde. Em seguida as moléculas de cDNA são inseridas em um vetor bacteriano de expressão e o vetor é transformado dentro de bactérias. As colônias bacterianas que contiverem o cDNA de insulina expressarão a proteína insulina. A proteína insulina é identificada por meio de sua

ligação com um anticorpo de insulina, conforme descrito anteriormente.

Detecção de clones específicos por complementação funcional

Em muitos casos, não temos uma sonda para o gene com a qual iniciar, mas temos uma mutação recessiva no gene de interesse. Esse gene pode ser mutante em uma bactéria ou levedura, ou até uma planta ou um camundongo. O objetivo dessa abordagem é identificar o clone que contém o gene de interesse, em virtude do fato de que ele restaurará a função eliminada pela mutação recessiva. Na prática, primeiramente é produzida uma biblioteca genômica ou de cDNA a partir de um organismo que apresenta o alelo do tipo selvagem do gene de interesse. O gene de interesse é um de milhares representados na biblioteca. Entretanto, apenas o gene de interesse apresenta a capacidade de complementar o organismo mutante e restaurar o fenótipo do tipo selvagem. Portanto, se formos capazes de introduzir a biblioteca na espécie que contém a mutação recessiva (ver seção 10.6), podemos detectar clones específicos na biblioteca por meio da sua capacidade de restaurar a função eliminada pela mutação recessiva. Esse procedimento é denominado **complementação funcional** ou **resgate de mutante**. O resumo geral do procedimento é como segue:

- Produza uma biblioteca contendo insertos de DNA doador recombinante do tipo selvagem a^+
- Transforme as células da linhagem celular mutante recessiva a^- com essa biblioteca de insertos de DNA
- Identifique os clones da biblioteca que produzem células transformadas com o fenótipo dominante a^+
- Recupere o gene a^+ do clone bacteriano ou de fago bem-sucedido.

Até agora, descrevemos as técnicas para transformar apenas células bacterianas. Você verá posteriormente nesse capítulo que o DNA pode ser introduzido em muitos organismos-modelo genéticos, incluindo *Saccharomyces cerevisiae* (levedura), *Caenorhabditis elegans* (verme nematódeo), *Arabidopsis thaliana* (planta) e *Mus musculus* (camundongo).

> CONCEITO-CHAVE Um gene clonado pode ser selecionado a partir de uma biblioteca por meio da utilização de sondas para descobrir a sequência de DNA do gene ou seu produto proteico ou por complementação de um fenótipo mutante.

Análise de DNA por Southern e Northern blot

Após ter amplificado seu produto de PCR ou selecionado um clone de interesse a partir de uma biblioteca genômica ou de cDNA, a próxima etapa é saber mais a respeito do DNA. Digamos que você tenha recuperado o cDNA da insulina a partir de um vetor de expressão e que deseja determinar os sítios de restrição na cópia genômica do gene da insulina. Talvez você deseje verificar se esses sítios diferem entre as diversas populações humanas. Você também pode desejar saber se o tamanho do mRNA da insulina varia entre as populações humanas. Alternativamente, você pode desejar determinar se um gene semelhante está presente no genoma de um organismo relacionado. Na seção a seguir, você verá que essas questões importantes podem ser respondidas por meio da utilização de técnicas relativamente simples. Nessas técnicas, misturas complexas de DNA ou RNA são ordenadas pelo tamanho e em seguida submetidas à hibridização com sondas para detectar moléculas de DNA relacionadas com alguma outra molécula de DNA.

O método mais extensivamente utilizado para a detecção de uma molécula dentro de uma mistura é o *blotting*, que tem início com a **eletroforese em gel** para separar as moléculas na mistura. Uma mistura de moléculas lineares de DNA é colocada em um poço formado em um gel de agarose. O gel é posicionado em uma caixa com eletrodos nas extremidades, de modo que os poços estejam na extremidade catódica (negativamente carregada) e o DNA, em virtude da sua carga negativa, migre até a extremidade anódica (positivamente carregada). A velocidade de migração das moléculas de DNA no gel é inversamente dependente do seu tamanho, tendo em vista que a agarose atua como uma peneira por meio da qual pequenas moléculas se movimentam mais fácil e rapidamente do que fragmentos maiores (Figura 10.14). Portanto, os fragmentos em classes de

tamanhos distintos formarão bandas distintas no gel. As bandas podem ser visualizadas por meio da coloração do DNA com brometo de etídio, que causa a fluorescência do DNA na luz ultravioleta. O tamanho absoluto de cada fragmento na mistura pode ser determinado por meio da comparação de sua distância de migração com um conjunto-padrão de fragmentos de tamanhos conhecidos. Se as bandas estiverem bem separadas, uma banda individual pode ser cortada do gel e a amostra de DNA pode ser purificada a partir da matriz de gel. Portanto, a eletroforese de DNA pode ser diagnóstica (demonstrando os tamanhos e as quantidades relativas dos fragmentos de DNA presentes) ou preparativa (útil no isolamento de fragmentos de DNA específicos).

O DNA genômico digerido por enzimas de restrição em geral produz tantos fragmentos que a eletroforese produz um esfregaço contínuo de DNA e nenhuma banda discreta. Uma sonda pode identificar um fragmento nessa mistura, com a utilização de uma técnica desenvolvida por E. M. Southern, denominada ***Southern blotting*** (Figura 10.15 A). Assim como a identificação de clones (ver Figura 10.12), essa técnica envolve a obtenção de uma impressão de moléculas de DNA em uma membrana utilizando-se a membrana para transferir o gel após a conclusão da eletroforese. O DNA deve ser primeiramente desnaturado, o que possibilita sua aderência à membrana. Em seguida a membrana é hibridizada com uma sonda marcada. Um autorradiograma ou uma fotografia das bandas fluorescentes revelará a presença de quaisquer bandas no gel que sejam complementares à sonda. Para detectar o gene da insulina, podemos aplicar esse protocolo ao DNA genômico humano digerido com enzimas de restrição na membrana, utilizando o cDNA da insulina como a sonda marcada.

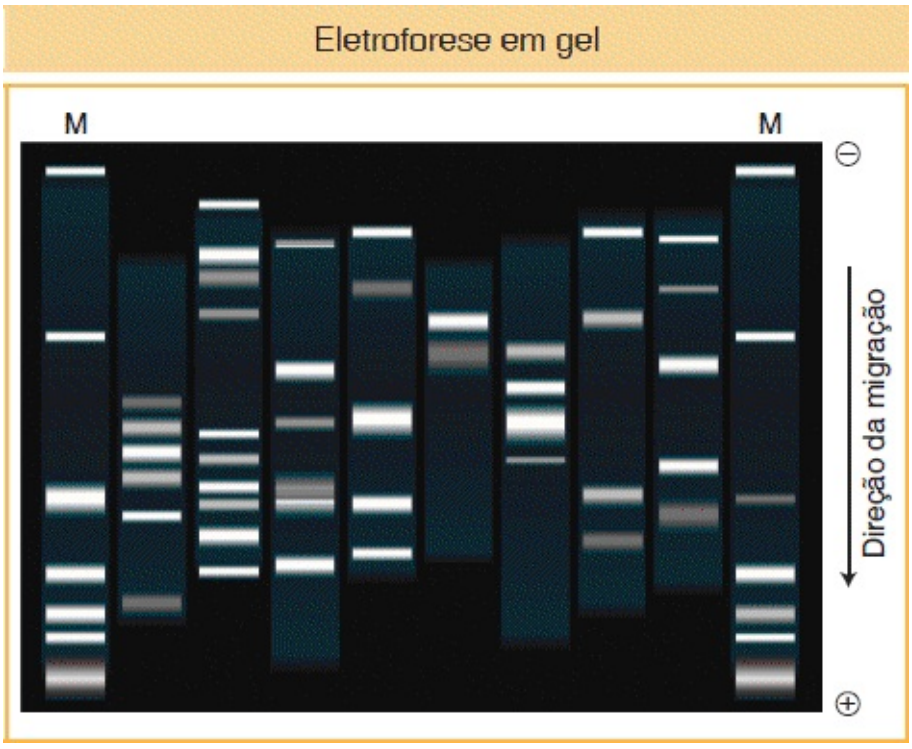


FIGURA 10.14 Misturas de fragmentos de DNA de tamanhos diferentes foram separadas eletroforeticamente em um gel de agarose. As amostras são oito vetores recombinantes tratados com *Eco*RI. As misturas são aplicadas em poços próximos da parte superior do gel e os fragmentos se movimentam a partir da extremidade negativa para a positiva, sob a influência de um campo elétrico, até diferentes posições, dependendo do tamanho. As bandas de DNA foram visualizadas por meio de coloração com brometo de etídio e fotografadas sob luz ultravioleta. (A letra M representa colunas que contêm fragmentos-padrão que atuam como marcadores para estimar o comprimento do DNA.) (Ingram Publishing/Thinkstock.)

A técnica de *Southern blotting* pode ser modificada para detectar uma molécula de RNA específica a partir de uma mistura de RNA fracionados em um gel. Essa técnica é denominada **RNA *blotting***, ou mais comumente, ***Northern blotting*** (graças ao senso de humor de algum cientista), para contrastá-la com a técnica de *Southern blotting* utilizada para a análise de DNA. O RNA separado por meio de eletroforese pode ser uma amostra do RNA total isolado a partir de um tecido ou a partir de um organismo inteiro. No exemplo demonstrado na Figura 10.15 B, fez-se a eletroforese em gel de RNA isolado das sementes de diversas plantas. Contrariamente ao DNA colocado em um gel, não há necessidade de digerir a amostra de RNA, tendo em vista que ela se encontra em moléculas de tamanho de transcritos discreto. Os géis de RNA são colocados

sobre uma membrana e hibridizados do mesmo modo como o DNA é hibridizado no *Southern blotting*. Uma aplicação da análise *Northern* é determinar se um gene específico é transcrito em um determinado tecido ou sob determinadas condições ambientais. Outra é determinar o tamanho do mRNA e se um RNA de tamanho semelhante pode ser detectado em plantas relacionadas de modo próximo (como na Figura 10.15 B).

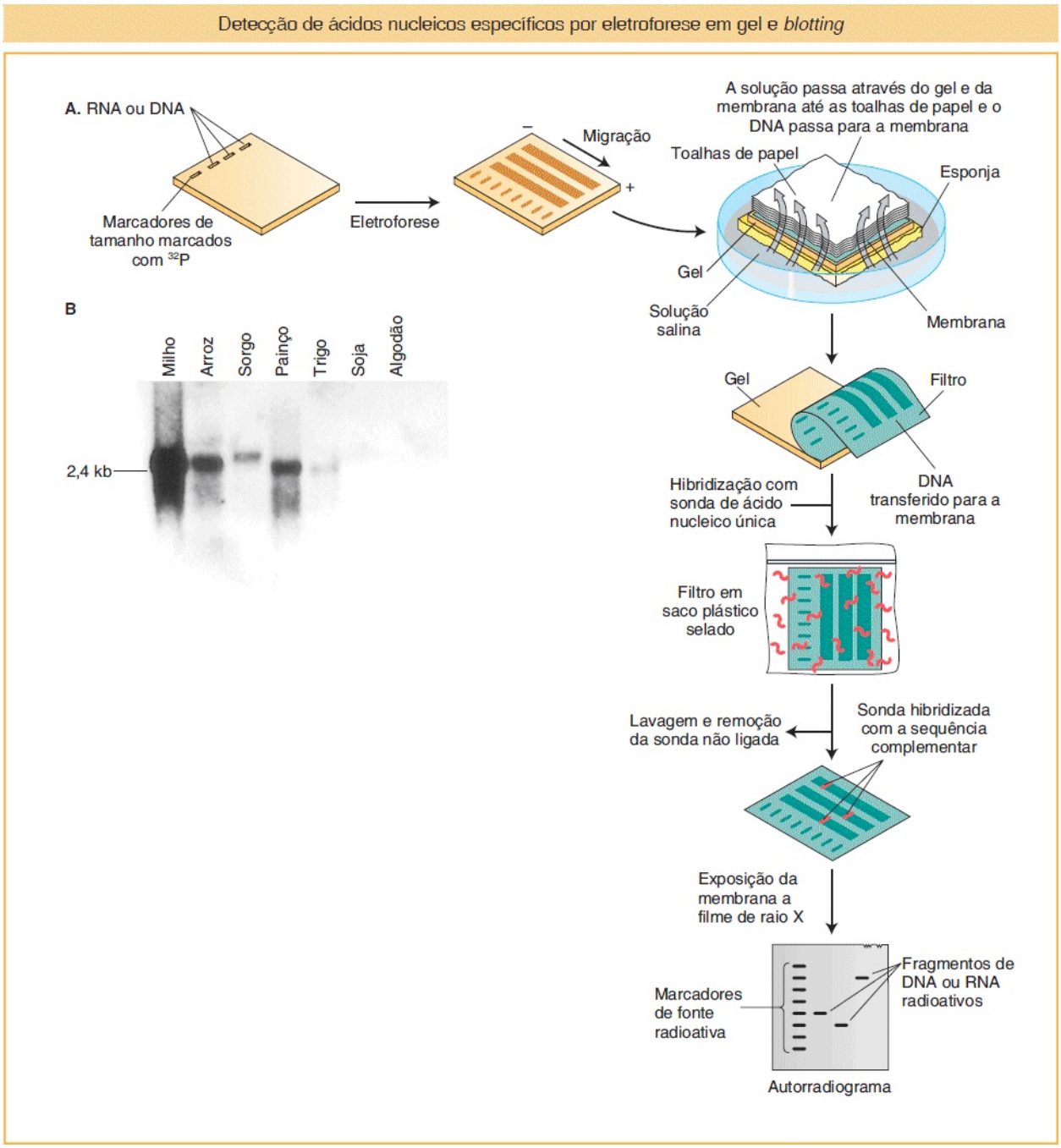


FIGURA 10.15 Neste exemplo, é utilizada uma sonda radioativa para identificar ácidos nucleicos específicos separados por meio de eletroforese em gel. **A.** Fragmentos de restrição de RNA ou DNA são aplicados em um gel de agarose e são submetidos à eletroforese. Os diversos fragmentos migram em diferentes velocidades, de acordo com seus respectivos tamanhos. O gel é colocado em tampão e recoberto por uma membrana e uma pilha de toalhas de papel. Os fragmentos são desnaturados até filamentos únicos, de modo que possam aderir à membrana. Eles são carregados para a membrana pelo tampão, que é absorvido pelas toalhas. Em seguida a membrana é removida e incubada com uma sonda unifilamentar marcada radioativamente, que é complementar à sequência-alvo. A sonda não ligada é lavada e removida, e um filme de raios X é exposto à

membrana. Tendo em vista que a sonda radioativa hibridizou apenas com seus fragmentos de restrição complementares, o filme será exposto apenas nas bandas que correspondem àqueles fragmentos. A comparação dessas bandas com marcadores revela o número e o tamanho dos fragmentos nos quais as sequências-alvo são encontradas. Esse procedimento é denominado *Southern blotting* quando o DNA é transferido para a membrana, e *Northern blotting* quando o RNA é transferido. **B.** *Northern blot* real, com RNA isolado a partir das sementes de diversas plantas. Uma única sonda de RNA é utilizada para identificar a presença de um *locus* único. Os resultados demonstram que o milho está relacionado de modo mais próximo ao arroz, ao sorgo e ao painço do que à soja ou ao algodão. (*B.* Susan Wessler.)

Iniciamos esta seção sobre a análise de *blot* indagando questões a respeito do gene da insulina e seu mRNA em populações humanas e em espécies correlatas. Com base nas técnicas anteriores, você consegue projetar experimentos de *Southern blot* e *Northern blot* para responder a essas questões? Você pode presumir que tem acesso às amostras de DNA e RNA genômicos necessários.

Portanto, vemos que o DNA clonado apresenta aplicação ampla como uma sonda utilizada para a detecção de um clone específico, um fragmento de DNA ou uma molécula de RNA. Em todos esses casos, observe que a técnica novamente explora a capacidade dos ácidos nucleicos com sequências de nucleotídios *complementares* de se encontrar e ligar uns aos outros.

> CONCEITO-CHAVE As técnicas de DNA recombinante que dependem da complementaridade com uma sonda de DNA clonado incluem sistemas de *blotting* e de hibridização para a identificação de clones específicos, fragmentos de restrição, ou mRNA para estimativa do tamanho de DNA ou RNA específicos.

Hibridização com sonda para uma proteína específica. Em relação às proteínas, a hibridização em geral é realizada por meio da utilização de anticorpos como sondas. Um anticorpo é uma proteína produzida pelo sistema imune de um animal, que se liga com alta afinidade a uma molécula tal como uma proteína específica (que atua como um *antígeno*), tendo em vista que o anticorpo apresenta um encaixe chave-fechadura específico com ela. Para a detecção da proteína, uma mistura proteica extraída de células é separada em bandas de proteínas distintas

por meio de eletroforese e em seguida é colocada sobre uma membrana (esse é um *Western blot*). A posição de uma proteína de interesse específica na membrana é revelada lavando-se membrana em uma solução de anticorpos obtidos de um coelho ou outro hospedeiro no qual o antígeno tenha sido injetado anteriormente. A posição da proteína é revelada pela posição da marcação que o anticorpo carrega.

10.4 Determinação da sequência de bases de um segmento de DNA

Após termos clonado e identificado nosso gene desejado ou termos amplificado esse gene por meio da PCR, tem início a tarefa de tentar compreender a sua função. A linguagem final do genoma é composta por filamentos dos nucleotídios A, T, C e G. A obtenção da sequência de nucleotídios completa de um segmento de DNA com frequência é uma parte importante da compreensão da organização de um gene e da sua regulação, sua relação com outros genes ou da função do RNA ou da proteína que ele codifica. De fato, a sequência de DNA pode ser utilizada para determinar a estrutura primária proteica, tendo em vista que, na maior parte, a tradução da sequência de ácidos nucleicos de uma molécula de cDNA para descobrir a sequência de aminoácidos de sua cadeia polipeptídica codificada é mais simples do que sequenciar diretamente o próprio polipeptídio. Nessa seção, consideramos as técnicas utilizadas para ler a sequência de nucleotídios do DNA.

Assim como as tecnologias do DNA recombinante e da PCR, o sequenciamento do DNA explora a complementaridade de pares de bases juntamente com uma compreensão da bioquímica básica da replicação do DNA. Foram desenvolvidas diversas técnicas, mas uma delas tem sido o método predominante utilizado até agora para sequenciar a maior parte das moléculas de DNA. Embora ainda seja a técnica mais comumente utilizada para sequenciar DNA mais curtos, novas tecnologias de sequenciamento suplantaram amplamente essa técnica quando o objetivo é determinar a sequências de genomas inteiros, conforme descrito em

detalhes no Capítulo 14.

Essa técnica de sequenciamento é denominada **sequenciamento didesóxi** ou, por vezes, **sequenciamento de Sanger**, em homenagem ao seu inventor. O termo *didesóxi* tem origem em um nucleotídio especial modificado, denominado didesoxinucleotídio trifosfato (genericamente, ddNTP). Esse nucleotídio modificado é a chave para a técnica de Sanger, em virtude da sua capacidade de bloquear a síntese contínua do DNA. Um didesoxinucleotídio não apresenta o grupo 3′-hidroxila, bem como o grupo 2′-hidroxila, que também está ausente em um desoxinucleotídio (Figura 10.16; compare com a Figura 7.5). Para que a síntese do DNA ocorra, a DNA polimerase deve catalisar uma reação entre o grupo 3′-hidroxila do último nucleotídio e o grupo 5′-fosfato do próximo nucleotídio a ser adicionado. Tendo em vista que um didesoxinucleotídio não apresenta o grupo 3′-hidroxila, essa reação não pode ocorrer e, portanto, a síntese do DNA é bloqueada no ponto da adição.

A lógica do sequenciamento didesóxi é direta. Suponha que desejamos ler a sequência de um segmento de DNA clonado de até 800 pares de bases. Esse segmento de DNA pode ser um inserto de plasmídio ou até mesmo um produto da PCR. Primeiramente, desnaturamos os dois filamentos desse segmento. Em seguida, criamos um *primer* para a síntese de DNA que hibridizará exatamente em um local no segmento de DNA clonado e, em seguida, adicionamos um “coquetel” especial de DNA polimerase, desoxinucleotídios trifosfato normais (dATP, dCTP, dGTP e dTTP) e uma pequena quantidade de um didesoxinucleotídio a uma das quatro bases (p. ex., ddATP). A polimerase começará a sintetizar o filamento de DNA complementar, iniciando a partir do *primer*, mas será interrompida em qualquer ponto no qual o didesoxinucleotídio trifosfato for incorporado na cadeia de DNA em crescimento em lugar do desoxinucleotídio trifosfato normal. Suponha que o segmento de DNA que estamos tentando sequenciar seja

5′ ACGGGATAGCTAATTGTTTACCGCCGGAGCCA 3′

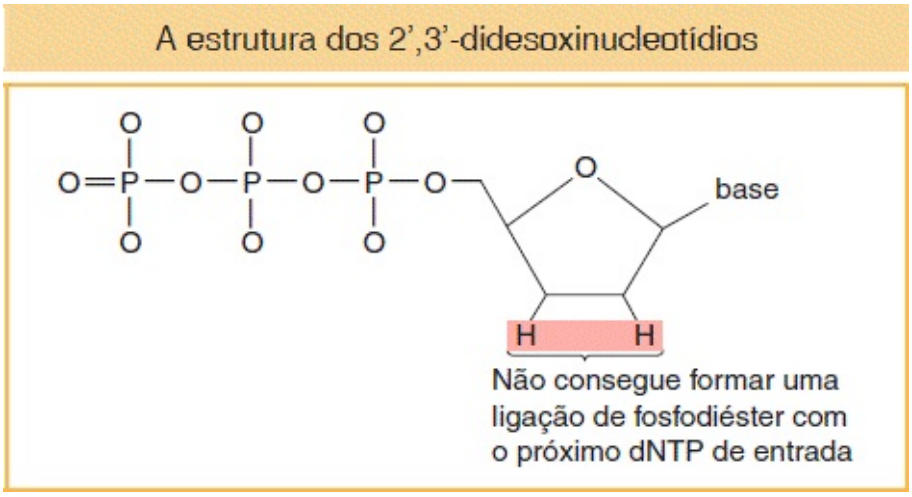


FIGURA 10.16 2′,3′-didesoxinucleotídios, que são empregados no método de sequenciamento de DNA de Sanger, estão ausentes no grupo hidroxila da ribose presente no DNA.

Em seguida, iniciaremos a síntese de DNA a partir de um *primer* complementar:

5′ ACGGGATAGCTAATTGTTTACCGCCGGAGCCA 3′
3′ CGGCCTCGGT 5′
← Sentido da síntese do DNA

Utilizando o coquetel especial para a síntese de DNA "enriquecido" com uma pequena quantidade de ddATP, por exemplo, criaremos um conjunto estabelecido de fragmentos de DNA que apresentam o mesmo ponto de início, mas diferentes pontos de término, tendo em vista que os fragmentos são interrompidos no ponto de inserção de ddATP, em vez de dATP, interrompendo a replicação do DNA. A série de diferentes cadeias de DNA interrompidas por ddATP se assemelha à lista de sequências a seguir. (*A indica o didesoxinucleotídio.)

Podemos produzir um arranjo dos referidos fragmentos para cada um dos quatro didesoxinucleotídios trifosfato possíveis em quatro coquetéis em separado (um enriquecido com ddATP, um com ddCTP, um com ddGTP e um com ddTTP). Cada um produzirá um arranjo diferente de fragmentos, com nenhum de dois coquetéis enriquecidos produzindo fragmentos do mesmo tamanho. Em seguida, os fragmentos de DNA gerados nos quatro coquetéis são separados e colocados em ordem com a utilização de eletroforese em gel. Ao aplicar os fragmentos em quatro colunas adjacentes de um gel de poliacrilamida que pode separar

fragmentos de DNA que variam em apenas um nucleotídio no comprimento, observamos que os fragmentos podem ser ordenados pelo comprimento, com os comprimentos aumentando em uma base por vez.

Os filamentos recém-sintetizados devem ser marcados de algum modo para tornar as bandas visíveis no gel. Os filamentos são marcados na medida em que são produzidos por meio da utilização de um *primer* marcado radioativamente ou quando houver um dos dNTP regulares com marcação radioativa. Também podem ser utilizadas marcações fluorescentes e, nesse caso, a fluorescência ocorre em cada ddNTP (ver a seguir).

Os produtos das referidas reações de sequenciamento didesóxi estão demonstrados na Figura 10.17. Tal resultado é uma escada de cadeias de DNA marcadas que aumentam em comprimento uma a uma e, assim, tudo o que precisamos fazer é ler o gel para ler a sequência de DNA do filamento sintetizado no sentido de 5′ para 3′.

Se a marcação for um corante fluorescente, um emissor de cor fluorescente diferente é utilizado para cada uma das quatro reações de ddNTP e um detector na extremidade do gel consegue distinguir cada uma das cores. As quatro reações ocorrem no mesmo tubo de ensaio, e os quatro conjuntos de cadeias de DNA formadas podem ser submetidos à eletroforese juntos. Portanto, quatro vezes mais sequências podem ser produzidas no mesmo período de tempo do que as que seriam produzidas por meio de reações em separado. Essa lógica é utilizada na detecção da fluorescência por máquinas de sequenciamento automático de DNA. Graças a essas máquinas, o sequenciamento de DNA pode prosseguir em um nível maciço e sequências de genomas inteiros foram obtidas por meio da ampliação dos procedimentos descritos nesta seção. A Figura 10.18 ilustra uma leitura de sequenciamento automático. Cada pico colorido representa um fragmento de DNA de tamanho diferente, que termina com uma base fluorescente detectada pelo *scanner* fluorescente do sequenciador automático de DNA; as quatro cores diferentes representam as quatro bases do DNA. As aplicações da tecnologia de sequenciamento automático em escala genômica são um foco importante do Capítulo 14.

5′	ATGGGATAGCTAATTGTTTACCGCCGGAGCCA	3′	Clone de DNA-molde
3′	CGGCCTCGGT	5′	*Primer* para a síntese
	←		Sentido da síntese de DNA
3′	*ATGGCGGCCTCGGT	5′	Fragmento didesóxi 1
3′	*AATGGCGGCCTCGGT	5′	Fragmento didesóxi 2
3′	*AAATGGCGGCCTCGGT	5′	Fragmento didesóxi 3
3′	*ACAAATGGCGGCCTCGGT	5′	Fragmento didesóxi 4
3′	*AACAAATGGCGGCCTCGGT	5′	Fragmento didesóxi 5
3′	*ATTAACAAATGGCGGCCTCGGT	5′	Fragmento didesóxi 6
3′	*ATCGATTAACAAATGGCGGCCTCGGT	5′	Fragmento didesóxi 7
3′	*ACCCTATCGATTAACAAATGGCGGCCTCGGT	5′	Fragmento didesóxi 8

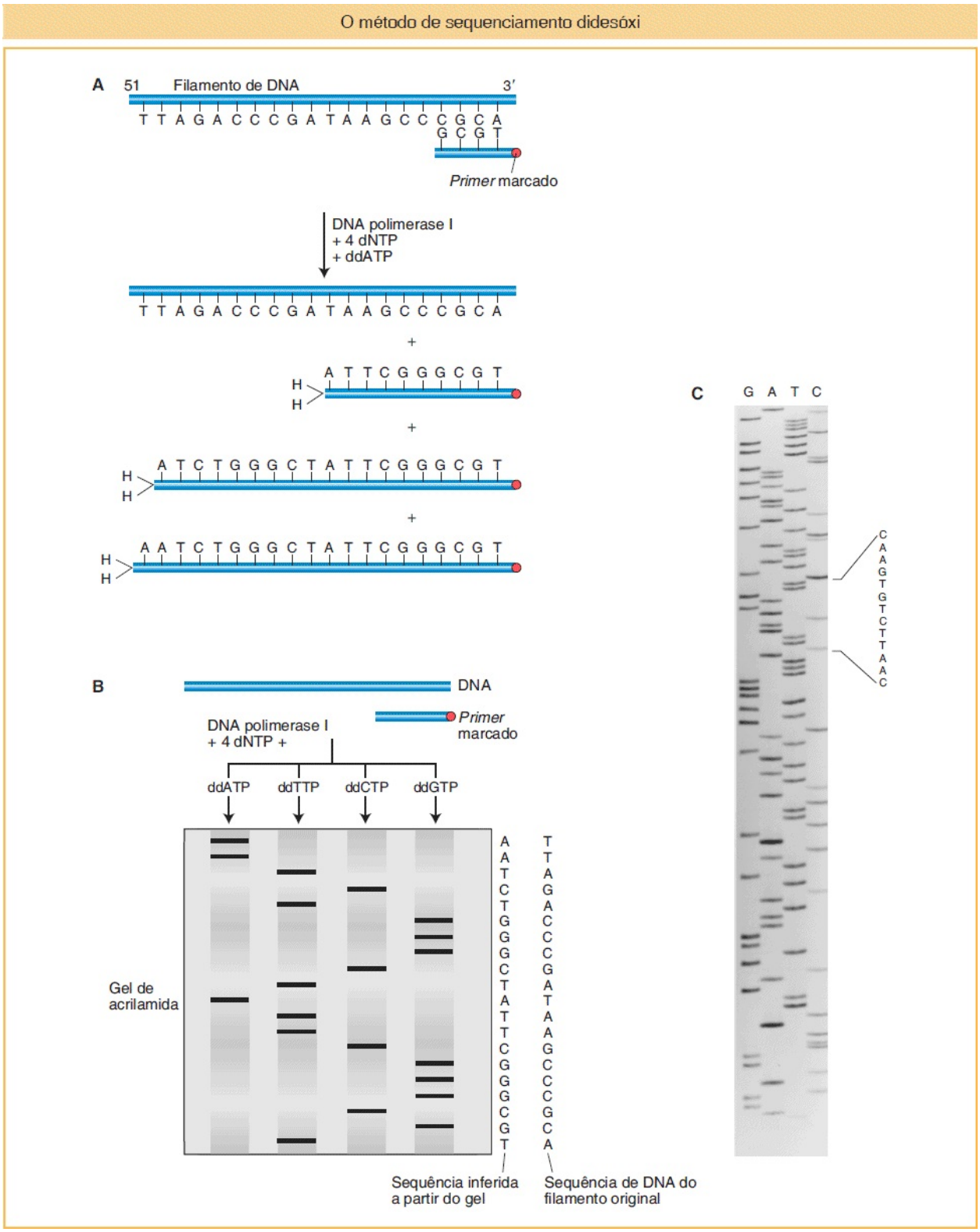


FIGURA 10.17 O DNA é sequenciado com eficiência por meio da inclusão de didesoxinucleotídios entre os nucleotídios utilizados para copiar um segmento de DNA. **A.** Um *primer* marcado (desenhado a partir da

sequência flanqueadora do vetor) é utilizado para iniciar a síntese de DNA. A adição de quatro didesoxinucleotídios diferentes (o ddATP está demonstrado aqui) interrompe a síntese aleatoriamente. **B.** Os fragmentos resultantes são separados eletroforeticamente e submetidos à autorradiografia. A sequência inferida está demonstrada à direita. **C.** Gel de sequenciamento de Sanger. (*C.* Loida Escote-Carlson, Ph.D.)

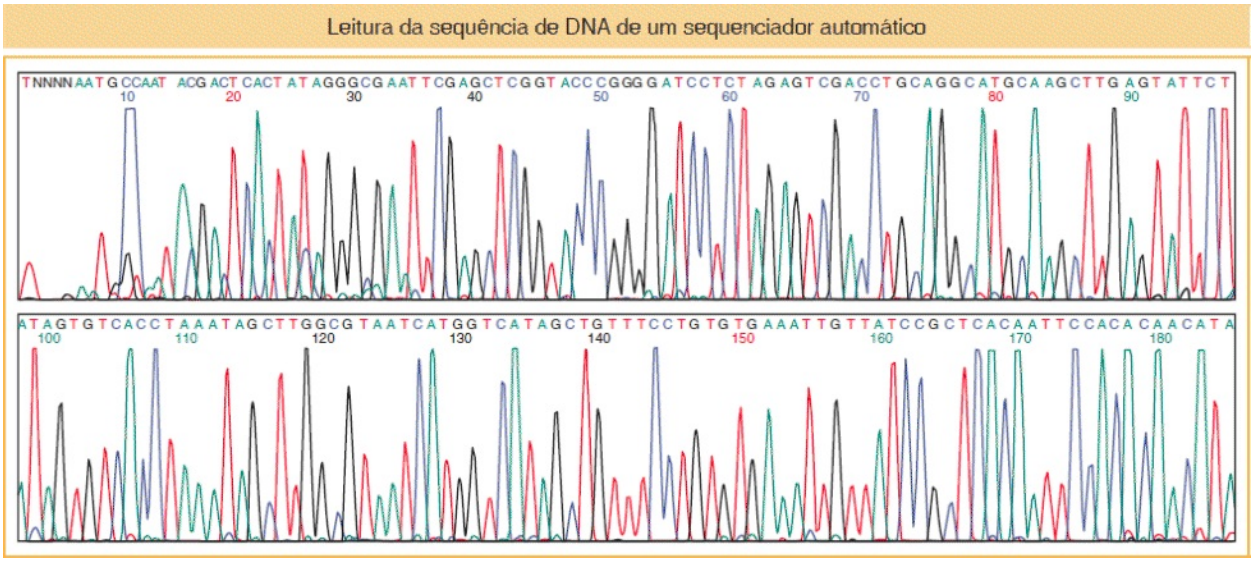


FIGURA 10.18 Cópia impressa de um sequenciador automático que utiliza corantes fluorescentes. Cada uma das quatro cores representa uma base diferente. A letra N representa uma base que não pode ser atribuída em virtude de os picos serem muito baixos. Observe que, se este fosse um gel, como na Figura 10.17 C, cada um destes picos corresponderia a uma das bandas escuras no gel; em outras palavras, estes picos coloridos representam uma leitura diferente do mesmo tipo de dados produzidos em um gel de sequenciamento.

> CONCEITO-CHAVE Um segmento de DNA clonado pode ser sequenciado por meio da caracterização das bases de término de um conjunto de fragmentos de DNA sintéticos truncados, cada um terminando em posições diferentes correspondentes à incorporação de um didesoxinucleotídio.

10.5 Alinhamento genético e mapas físicos para isolar genes específicos

Antes que sequências genômicas completas estivessem disponíveis, a clonagem molecular de genes em relação a distúrbios genéticos tais como a fibrose cística (FC) e determinados cânceres era uma tarefa árdua. A identificação de genes

clonados em bibliotecas genômicas com frequência necessitava de um projeto de pesquisa maior, envolvendo os esforços cooperativos de diversos laboratórios. O processo é conhecido como **clonagem posicional** e sua estratégia é utilizar a posição genética para isolar o gene subjacente ao traço. Até mesmo com a disponibilidade de sequências de genomas inteiros, com frequência ainda é necessário primeiramente mapear traços que não foram associados a um produto gênico antes de identificar o gene responsável pela atividade.

Para iniciar a clonagem posicional, pesquisadores primeiramente precisam mapear o gene responsável por um traço em particular. Para mapear o gene, os pesquisadores podem testar a ligação com marcadores de localização conhecida, conforme descrito no Capítulo 4. Os marcadores podem ser RFLP (polimorfismos de comprimento de fragmento de restrição), SNP (polimorfismos de nucleotídio único), ou outros polimorfismos moleculares (ver Capítulos 4 e 14), ou podem ser pontos de quebra cromossômica bem-mapeados (ver Capítulo 17). Marcadores em cada lado do gene de interesse são os melhores, tendo em vista que delimitam a possível localização daquele gene.

É importante ter em mente que o mapeamento de um gene por meio de qualquer um desses procedimentos serve apenas para localizar a "vizinhança" cromossômica do gene; nenhum desses procedimentos identifica diretamente o gene. Portanto, as regiões delimitadas pelos marcadores moleculares tais como SNP e RFLP normalmente contêm muitos genes espalhados por centenas de milhares ou mesmo milhões de pares de bases. Para identificar o gene correto responsável por um traço em particular, os pesquisadores devem ser capazes de analisar toda a vizinhança em relação ao gene de interesse. Em organismos-modelo para os quais a sequência de todo o genoma está disponível, a vizinhança local, com seus diversos genes, pode ser obtida simplesmente a partir de uma base de dados em computador (ver Capítulo 14). A partir da análise da sequência desses genes da "vizinhança", são escolhidos os candidatos mais prováveis que possam representar o gene que está sendo buscado.

A seguir, discutiremos brevemente como a clonagem posicional foi utilizada antes da disponibilidade da sequência genômica humana para isolar o gene responsável pela devastadora doença humana fibrose cística (o gene da FC).

Embora essa técnica não seja mais necessária para a identificação de genes humanos, ela ainda é utilizada para identificar genes específicos em organismos nos quais uma sequência genômica ainda deva ser determinada.

Utilização de clonagem posicional para identificar o gene de uma doença humana

Seguiremos os métodos utilizados para identificar a sequência genômica do gene da fibrose cística. Nenhum defeito bioquímico primário era conhecido na ocasião em que o gene foi isolado e, assim, tratava-se muito mais um gene em busca de uma função.

Triagens genéticas dependentes do mapeamento gênico podem ser utilizadas para dissecar qualquer processo biológico. Entretanto, triagens genéticas não podem ser utilizadas em seres humanos, tendo em vista que não podemos criar intencionalmente mutantes humanos. Em vez disso, são realizadas análises de heredogramas de grandes famílias com o traço da doença (ver Capítulo 2), quando as referidas informações estão disponíveis, para determinar a posição do defeito genético que causa uma doença tal como a fibrose cística. Observa-se que membros de uma família que têm a doença apresentam um ou mais marcadores moleculares (anormalidades cromossômicas, traços ligados de localização conhecida, SNP e assim por diante) em comum, que não são observados em outras famílias (ver uma discussão sobre os marcadores moleculares no Capítulo 4). A ligação a marcadores moleculares em famílias com FC localizou o gene da FC no braço longo do cromossomo 7, entre as bandas 7q22 e 7q31.1. Acreditava-se que o gene da FC estivesse nessa região, mas entre esses marcadores há 1,5 centimorgan (unidades de mapa) de cromossomo, um vasto terreno inexplorado de mais de 1 milhão de bases. Para chegar mais perto, era necessário gerar mais marcadores moleculares nessa região. O método geral para isolamento de marcadores moleculares (descrito no Capítulo 4) é identificar uma região do DNA que seja polimórfica em indivíduos ou populações que diferem em relação ao traço de interesse. Ao encontrar marcadores moleculares adicionais ligados ao gene da FC, os geneticistas estreitaram a região que contém o gene da FC para aproximadamente 500 kbp, ainda uma distância considerável.

Foi criado um mapa físico dessa região inteira; ou seja, um conjunto aleatório de clones dessa região foi posicionado na ordem correta. Isso foi parcialmente realizado por meio da utilização de uma técnica denominada **caminhada cromossômica** (*chromosome walk*) (Figura 10.19). A ideia básica é utilizar a sequência do marcador próximo como uma sonda para identificar um segundo conjunto de clones que se sobrepõe ao clone do marcador que contém o marco, mas que se estende para fora a partir dele em uma de duas direções (em direção ao alvo ou para longe do alvo). Fragmentos terminais dos novos conjuntos de clones podem ser utilizados como sondas para a identificação de um terceiro conjunto de clones sobrepostos a partir da biblioteca genômica. Desse modo tedioso, geneticistas identificaram o clone que contém os marcadores moleculares que estão mais firmemente ligados ao traço da FC e sequenciaram aquele clone. Com a sequência em mãos, a caçada por quaisquer genes ao longo desse trecho de DNA, que contivessem os genes e as sequências não codificadoras, pôde ser iniciada.

No exemplo da FC, os genes candidatos foram identificados por meio da observação de características, tais como sinais de início e parada, comuns aos genes. As sequências dos genes candidatos e dos cDNA em seguida foram comparadas entre indivíduos normais e pacientes com FC. Foi observada uma mutação em um gene candidato que aparecia em todos os pacientes com FC analisados, mas em nenhum dos indivíduos normais. Essa mutação era uma deleção de três pares de bases, que eliminava uma fenilalanina da proteína. Por sua vez, a partir da sequência inferida, foi prevista a estrutura tridimensional da proteína. Essa proteína é estruturalmente semelhante às proteínas de transporte de íons em outros sistemas, sugerindo que um defeito de transporte é a causa primária da FC. Quando utilizado para transformar linhagens celulares mutantes de pacientes com FC, o gene do tipo selvagem restaurou a função normal; esse "resgate" fenotípico foi a confirmação final de que a sequência isolada era de fato o gene da FC.

Outros genes humanos isolados por meio de clonagem posicional incluem aqueles envolvidos em diversas doenças hereditárias, incluindo doença de Huntington, câncer de mama, síndrome de Werner (ver Capítulo 7) e

suscetibilidade à asma. Em virtude da facilidade relativa de cruzamento de plantas, a clonagem posicional tem sido uma técnica muito poderosa para isolar genes envolvidos em muitos processos, incluindo a identificação de genes que contribuíram para a domesticação de cultivos.

Utilização do mapeamento fino para identificar genes

Atualmente, o processo extremamente tedioso de "caminhada" até o gene deixou de ser necessário para qualquer organismo em relação ao qual uma sequência genômica esteja disponível. Pesquisadores ainda iniciam a caçada ao gene por meio da identificação de dois marcadores moleculares que flanqueiem de modo próximo o gene de interesse (ver Figura 10.19). Na Figura 10.20 A, os dois marcadores flanqueadores iniciais estão indicados como marcador de início 1 e marcador de início 2. No intervalo entre os dois marcadores de início, há sete genes cuja existência é conhecida (genes A a G). Qual desses genes é responsável pelo traço de interesse?

Os pesquisadores tentam estreitar o intervalo que encerra o gene de interesse. Para isso, selecionam marcadores adicionais, localizados entre os marcadores de início, a partir de bancos de dados *on-line* de sequências genômicas e de marcadores. Eles devem selecionar marcadores que apresentem um alelo em indivíduos com o traço fenotípico e um alelo diferente em indivíduos sem o fenótipo. Esses marcadores adicionais são denominados marcadores de finalização; na Figura 10.20 A, eles são os marcadores de finalização 1 a 7. O objetivo é encontrar os marcadores mais estreitamente ligados ao gene de interesse.

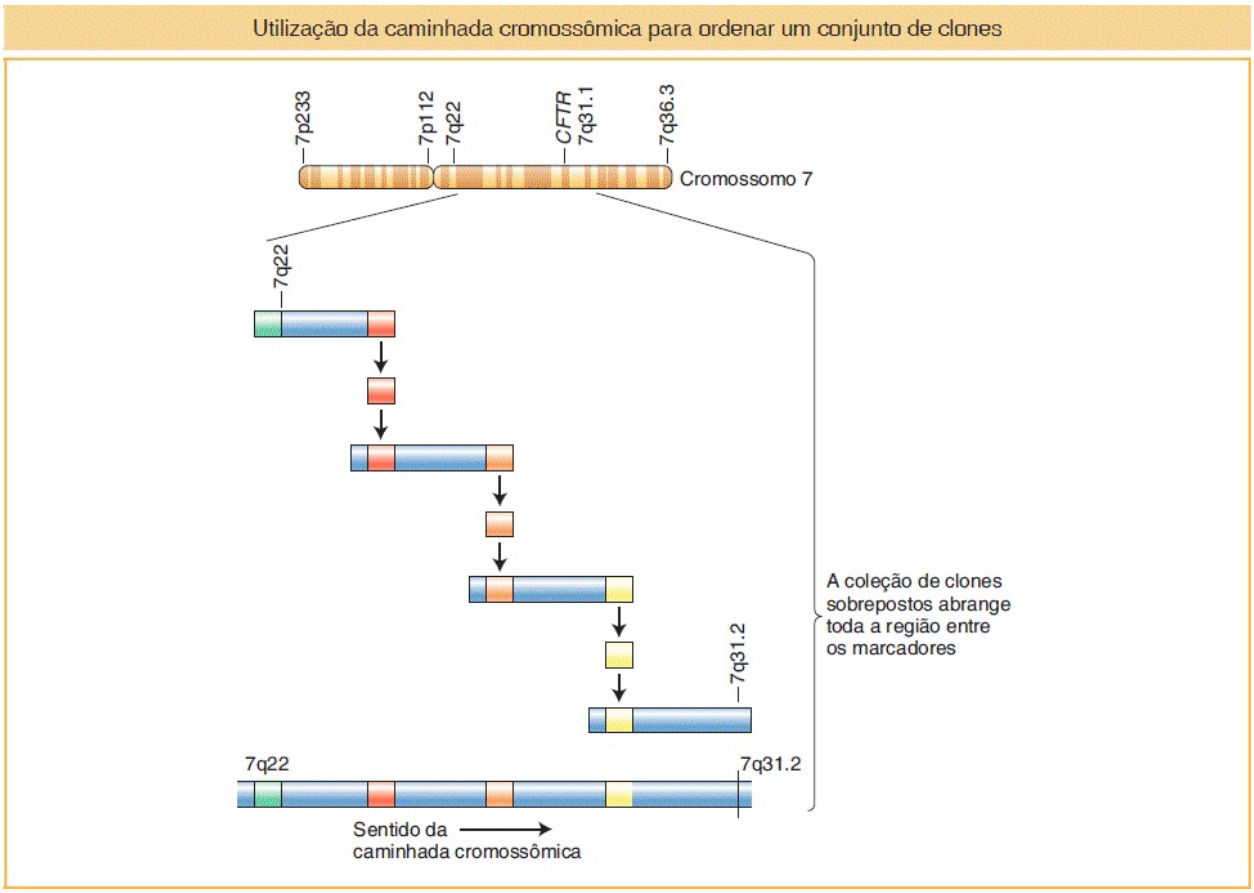


FIGURA 10.19 Esta caminhada cromossômica tem início com um fago recombinante ou clone BAC obtido a partir de uma biblioteca que contém grandes insertos que representam um genoma eucariótico inteiro. No exemplo demonstrado, o marcador molecular 7q22 é utilizado como sonda em uma biblioteca genômica humana. Apenas os insertos de DNA estão demonstrados. O inserto de DNA selecionado pela sonda é utilizado em seguida para isolar outro fago recombinante ou BAC que contenha um segmento adjacente de DNA eucariótico. Esta caminhada ilustra como iniciar no marcador molecular 7q22 e chegar ao marcador 7q31.1, que está do outro lado do gene da FC.

A próxima etapa é procurar indivíduos nos quais um evento de *crossover* raro tenha ocorrido dentro da região limitada pelos dois marcadores de início. Normalmente, os marcadores de início estão a uma distância de 1 ou mais centimorgans. Como tal, haverá, em média, um único *crossover* em um intervalo desse tamanho em uma progênie de 100. Em virtude do grande número de progênie necessária, esse tipo de mapeamento tem sido aplicado com sucesso em muitos organismos-modelo genéticos (tais como a mosca-das-frutas e *C. elegans*), mas é especialmente de sucesso em plantas, tendo em vista que os cruzamentos de

plantas podem produzir centenas ou milhares de progênies, que podem ser obtidas e analisadas. A Figura 10.21 demonstra a série de cruzamentos utilizados para produzir progênie contendo recombinantes entre os marcadores de início M1 e M2 (ver Figura 10.20).

Os alelos marcadores revelam se os marcadores e os genes na mesma região foram herdados da mãe ou do pai. Nos recombinantes, parte do segmento entre os marcadores de início em um cromossomo terá sido herdada da mãe e uma parte do pai (Genitor 2 na Figura 10.21). Na Figura 10.20, o traço é uma doença herdada do pai. Em indivíduos com a doença, o *crossover* criou uma região, colorida em azul na Figura 10.20 B, que é homozigota em relação ao gene da doença. Portanto, o gene deve estar localizado na região azul em todos os indivíduos com a doença. Comparando todos os indivíduos com a doença, você pode verificar que apenas a região azul compartilhada por todos é a região na qual está localizado o gene *D*.

Em outras palavras, *D* é o gene presente em regiões do genoma que são homozigotas em relação ao alelo paterno em indivíduos com a doença. Se o número de indivíduos nos heredogramas ou nas populações de estudo for suficientemente grande, então pode ser possível identificar não apenas o gene em questão, mas também a lesão da doença, que é o sítio polimórfico no gene que controla a diferença no traço. Observe que esse processo não envolve a clonagem de segmentos de DNA nas bibliotecas de BAC ou a triagem das referidas bibliotecas. Como tal, ele é mais adequadamente denominado **mapeamento fino**, em vez de clonagem posicional.

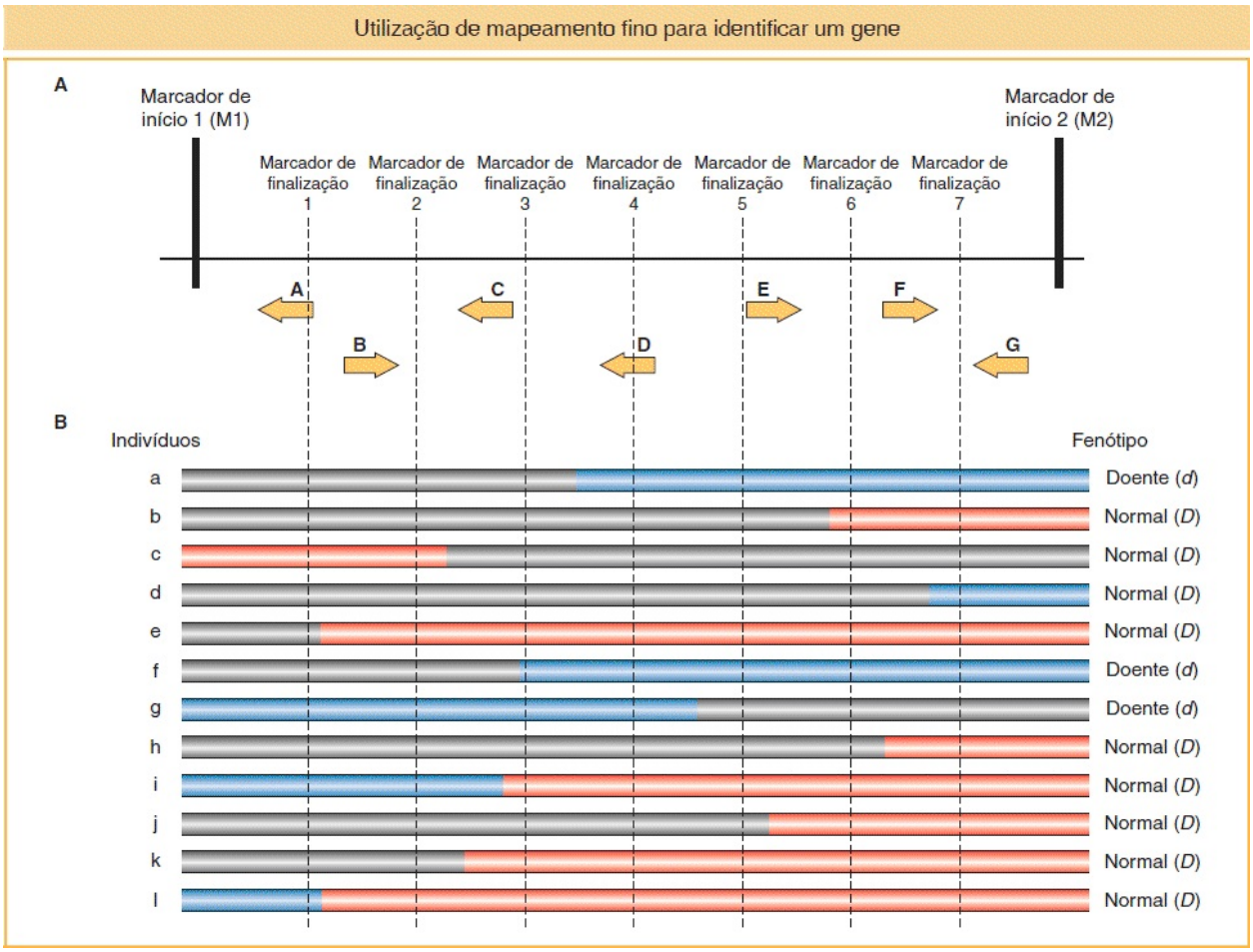


FIGURA 10.20 Uma caçada moderna de um gene utiliza marcadores e sequências obtidos a partir de bancos de dados *on-line* para determinar os genótipos em uma região de DNA em relação a grandes números de indivíduos com e sem o traço da doença. Os indivíduos demonstrados aqui foram derivados da progênie F_2 na Figura 10.21. O gene-alvo é o alelo gênico compartilhado por todos com a doença, o gene *D*. *Vermelho* é homozigoto para o alelo *D* dominante do Genitor 1 (normal); *azul* é homozigoto para o alelo *d* recessivo do Genitor 2 (doente, mutante); *cinza* é heterozigoto.

Os investigadores ainda devem trabalhar muito e superar muitos obstáculos para isolar os genes que controlam doenças ou outros traços. Primeiramente, eles precisam ter grandes amostras de indivíduos para assegurar que possam identificar eventos de *crossover* raros entre todos os genes. Tipicamente, isso significa dispor de milhares de indivíduos. Sem grandes populações, os investigadores podem recuperar *crossovers* apenas a cada conjunto de diversos genes e, consequentemente, não teriam certeza sobre qual desses seria o gene causador. Por exemplo, com apenas os indivíduos *a* e *b* na Figura 10.20 B, um

investigador somente poderia estreitar a busca a quatro genes (*D, E, F, G*). Em segundo lugar, embora os bancos de dados *on-line* contenham listas de marcadores com base no DNA, tais como SNP, nem todos apresentam alelos que estejam presentes somente em indivíduos com o traço de interesse em um heredograma ou cruzamento em particular, de modo que os pesquisadores devem primeiramente triar um grande número de marcadores para encontrar aqueles que o apresentam. Finalmente, os investigadores precisam determinar a sequência completa do DNA do alelo do gene da doença para identificar a lesão causadora. Na maior parte dos casos, a sequência genômica *on-line* conterá o alelo do tipo selvagem. O alelo da doença é mais facilmente sequenciado com a utilização da PCR para amplificar aquele alelo do DNA de indivíduos afetados e sem de fato clonar o DNA. A mutação precisa pode então ser deduzida a partir da sequência de DNA do produto da PCR.

> CONCEITO-CHAVE Até mesmo com o acesso às sequências de genomas inteiros, o isolamento dos genes defeituosos que causam doenças tem início com o mapeamento genético do traço da doença. Dispondo de marcadores estreitamente ligados flanqueando o traço, os investigadores podem utilizar o mapeamento fino para estreitar a busca do gene de interesse.

As seções precedentes introduziram as técnicas fundamentais que revolucionaram a genética. A seção final desse capítulo enfocará a aplicação dessas técnicas na engenharia genética.

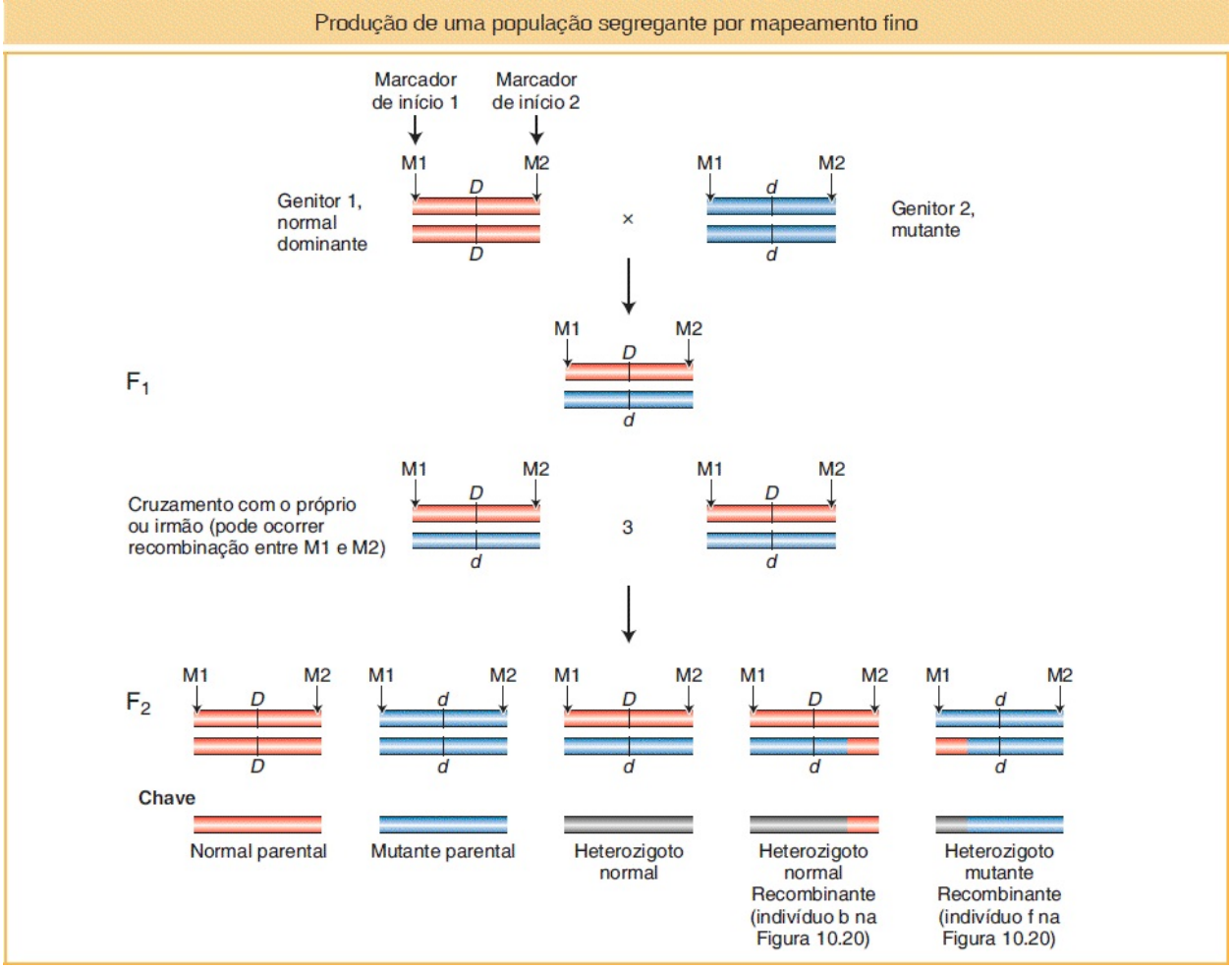


FIGURA 10.21 Uma caçada moderna de um gene com frequência tem início com um cruzamento entre genitores com traços contrastantes. No exemplo demonstrado, o Genitor 1 carrega o alelo do tipo selvagem (*D*) e é normal, enquanto o Genitor 2 carrega o alelo mutante (*d*) e, como resultado, apresenta a doença. Toda a progênie F_1 é heterozigota em todos os *loci* e é normal. A progênie F_2 está segregando em relação à doença. A maior parte da progênie com a doença apresenta o genótipo do Genitor 2 (mutante parental). Indivíduos mutantes raros sofreram um evento de recombinação em um dos genitores no intervalo cromossômico entre os marcadores M1 e M2 (mutante heterozigoto, parte inferior à direita).

10.6 Engenharia genética

Graças à tecnologia do DNA recombinante, os genes podem ser isolados e caracterizados como sequências de nucleotídios específicas. Mas até mesmo essa conquista não é o fim da história. Veremos em seguida que o conhecimento de uma sequência com frequência é o início de uma nova rodada de manipulação

genética. Quando caracterizada, uma sequência pode ser manipulada para alterar o genótipo de um organismo. A introdução de um gene alterado em um organismo tornou-se central para a pesquisa genética básica, mas também encontra ampla aplicação comercial. Dois exemplos da última são (1) cabras que secretam em seu leite antibióticos derivados de um fungo e (2) plantas protegidas do congelamento por meio da incorporação de genes "anticongelamento" de um peixe ártico em seus genomas. A utilização de técnicas de DNA recombinante para alterar o genótipo e o fenótipo de um organismo é denominada *engenharia genética*, e a sua aplicação prática é denominada *biotecnologia*.

As técnicas de engenharia genética descritas na primeira parte deste capítulo foram originalmente desenvolvidas em bactérias. Portanto, essas técnicas precisaram ser estendidas para eucariotos-modelo, que constituem uma grande proporção dos organismos-modelo de pesquisas. Os genes eucarióticos ainda são tipicamente clonados e sequenciados em hospedeiros bacterianos, mas eventualmente são introduzidos em um eucarioto, seja a espécie doadora original, ou uma completamente diferente. O gene transferido é denominado **transgene**, e o produto modificado é denominado **organismo transgênico**.

O transgene pode ser introduzido em uma célula eucariótica por meio de uma diversidade de técnicas, incluindo transformação, injeção, infecção bacteriana ou viral e bombardeamento com partículas de tungstênio ou ouro revestidas por DNA com a utilização de um disparador de genes (Figura 10.22). Quando o transgene entra em uma célula, ele se movimenta até o núcleo, onde deve se tornar uma parte estável do genoma por meio da inserção em um cromossomo ou (apenas em algumas poucas espécies) da replicação como parte de um plasmídio. Se a inserção ocorre, o transgene pode substituir o gene residente por meio de *recombinação homóloga* ou inserir-se **ectopicamente** — ou seja, em outros locais no genoma. Transgenes de outras espécies tipicamente se inserem ectopicamente.

> CONCEITO-CHAVE A transgênese pode introduzir material genético novo ou modificado em células eucarióticas.

Agora nos voltamos para alguns exemplos em fungos, plantas e animais.

Engenharia genética em Saccharomyces cerevisiae

É justo dizer que o *S. cerevisiae* é o modelo genético eucariótico facilmente manipulado mais sofisticado. A maior parte das técnicas tipicamente utilizada para a engenharia genética eucariótica foi desenvolvida em leveduras; assim, consideraremos as vias gerais em relação à transgênese em leveduras.

Os vetores de leveduras mais simples são os plasmídios integrativos (YIp), derivados de plasmídios bacterianos nos quais o DNA de levedura de interesse foi inserido. Quando transformados em células de levedura, esses plasmídios se inserem nos cromossomos da levedura, em geral por meio de recombinação homóloga com o gene residente, seja por meio de um *crossover* único ou duplo (Figura 10.23). Como resultado, o plasmídio inteiro é inserido, ou o alelo-alvo é substituído pelo alelo no plasmídio. O último é um exemplo de *substituição gênica* — nesse caso, a substituição do gene original na célula de levedura por um gene modificado por engenharia genética. A substituição gênica pode ser utilizada para deletar um gene ou substituir um alelo mutante pelo seu correspondente do tipo selvagem ou, contrariamente, substituir um alelo do tipo selvagem por um mutante. As referidas substituições podem ser detectadas por meio do plaqueamento de células em um meio seletivo para um alelo marcador no plasmídio.

A origem bacteriana da replicação é diferente das origens eucarióticas e, assim, os plasmídios bacterianos não replicam em leveduras. Portanto, o único modo pelo qual os referidos vetores podem gerar um genótipo modificado estável é se eles estiverem integrados ao cromossomo da levedura.

Engenharia genética em plantas

A tecnologia do DNA recombinante introduziu uma nova dimensão ao esforço de desenvolver variedades de cultivos melhores. A diversidade genética deixou de ser conquistada somente por meio da seleção de variantes de uma determinada espécie. Atualmente, pode-se introduzir o DNA de outras espécies de plantas, animais, ou até mesmo bactérias, produzindo **organismos geneticamente**

modificados (OGM). As modificações do genoma que se tornaram possíveis por meio dessa tecnologia são quase ilimitadas. Em resposta às novas possibilidades, uma parcela do público expressou a preocupação de que a introdução de OGM nos suprimentos alimentares possa produzir problemas de saúde inesperados. A preocupação a respeito dos OGM é uma faceta de um debate público em andamento a respeito de complexas questões de saúde pública, segurança, ética e questões educacionais originadas pelas novas tecnologias genéticas.

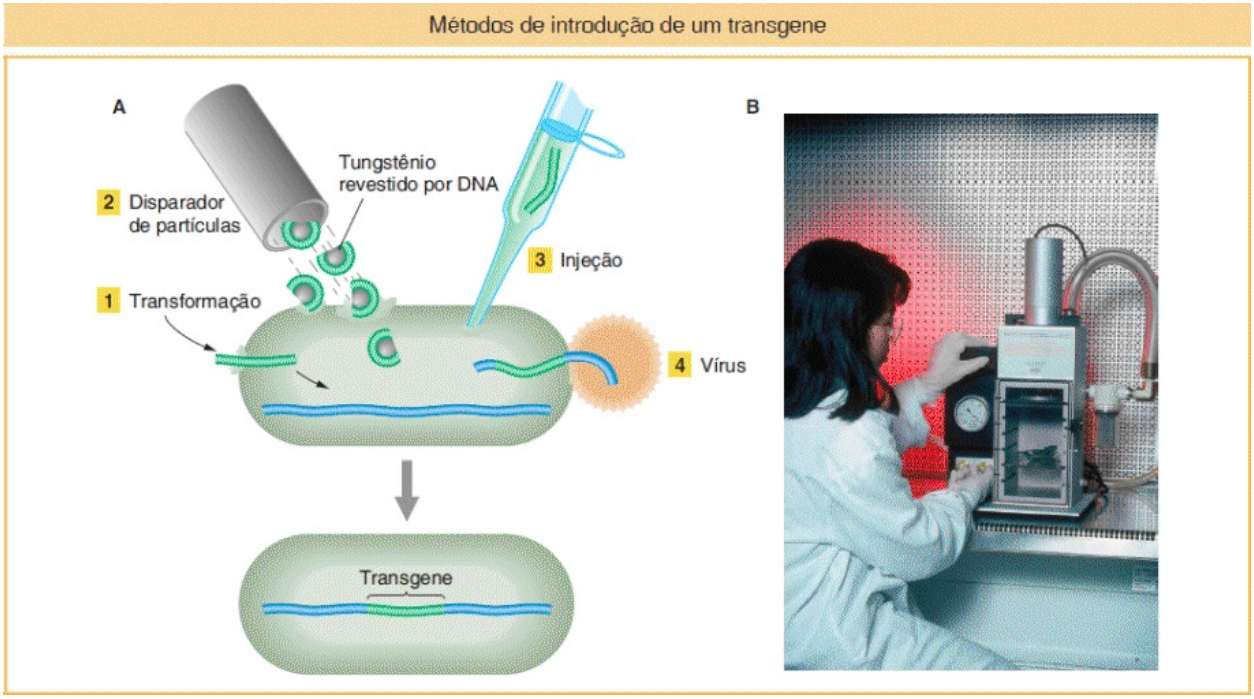


FIGURA 10.22 A. Quatro modos diferentes de introduzir um DNA exógeno em uma célula. B. Um disparador de genes. (*B.* Matt Meadows/Getty Images.)

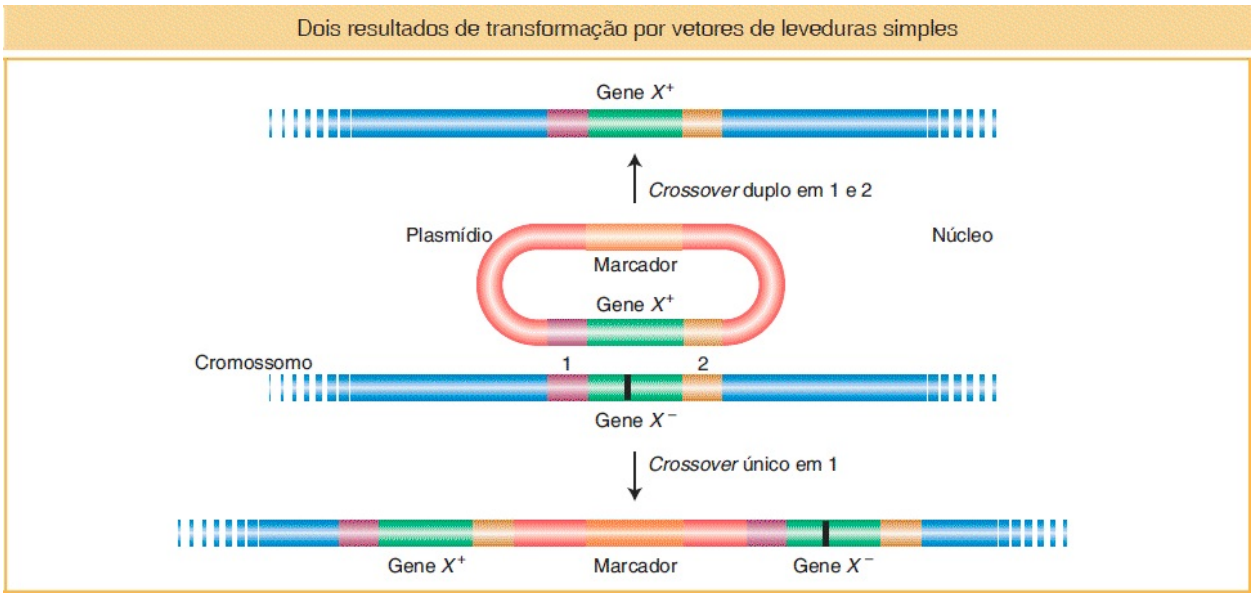


FIGURA 10.23 Um plasmídio que contém um alelo ativo (gene X^+) se insere em uma linhagem de levedura receptora que contém um gene defeituoso (X^-) por meio de recombinação homóloga. O resultado pode ser a substituição do gene defeituoso X^- pelo X^+ (*parte superior*) ou a sua retenção juntamente com o novo alelo (*parte inferior*). O sítio mutante do gene X^- é representado como uma barra preta vertical. *Crossovers* únicos na posição 2 também são possíveis, mas não estão demonstrados.

Um vetor rotineiramente utilizado para produzir plantas transgênicas é derivado do **plasmídio Ti**, um plasmídio natural de uma bactéria do solo denominada *Agrobacterium tumefaciens*. Essa bactéria causa o que é conhecido como *doença galha da coroa*, na qual a planta produz crescimentos descontrolados, denominados tumores ou galhas. A base para a produção de tumor é um grande (200 kb) plasmídio de DNA circular — o *plasmídio Ti (indutor de tumor)*. Quando a bactéria infecta uma célula da planta, uma parte do plasmídio Ti é transferida e inserida, aparentemente mais ou menos de modo aleatório, no genoma da planta hospedeira (Figura 10.24). A região do plasmídio Ti que se insere na planta hospedeira é denominada *T-DNA*, em referência ao DNA de transferência. Os genes cujos produtos catalisam essa transferência do T-DNA estão localizados em uma região do plasmídio Ti separada da própria região do T-DNA.

O comportamento natural do plasmídio Ti torna-o bem-adequado para o papel de um vetor para a engenharia genética em plantas. Em particular, qualquer DNA

que seja inserido entre as sequências de bordas de T-DNA (extremidades de 24 pb) pode ser mobilizado por outras funções proporcionadas pelo plasmídio Ti e inserido em cromossomos de plantas. Portanto, os cientistas foram capazes de eliminar toda a sequência de T-DNA entre as bordas (incluindo os genes que causam tumores) e substituí-la pelo(s) gene(s) de interesse e por um marcador selecionável (p. ex., resistência à canamicina). Um método para introduzir o T-DNA no genoma da planta está demonstrado na Figura 10.25. Bactérias que contêm esse T-DNA e outras que contêm T-DNA similares modificados por engenharia genética são utilizadas para infectar segmentos cortados de tecido vegetal, tais como discos de folhas perfuradas. Se os discos de folhas forem colocados em um meio contendo canamicina, apenas as células da planta que adquiriram o gene kan^R modificado por engenharia genética no T-DNA sofrerão divisão celular. As células transformadas crescem em um aglomerado, ou calo, que pode ser induzido a formar brotos e raízes. Esses calos são transferidos para o solo, onde podem se desenvolver em plantas transgênicas. Tipicamente, apenas uma cópia única da região do T-DNA se insere no genoma de uma determinada planta, onde ele se segrega na meiose como um alelo mendeliano regular (Figura 10.26). A presença do inserto pode ser verificada por meio da triagem do tecido transgênico em relação a marcadores genéticos transgênicos ou por meio da triagem do DNA purificado com uma sonda de T-DNA em uma hibridização por *Southern*.

Plantas transgênicas que carregam qualquer um de uma diversidade de genes exógenos estão em uso atualmente, incluindo plantas de cultivo que carregam genes que conferem resistência a determinadas pragas bacterianas ou fúngicas, e muitas mais estão em desenvolvimento. Não apenas as qualidades das próprias plantas estão sendo manipuladas, mas, assim como os microrganismos, as plantas também estão sendo utilizadas como “fábricas” convenientes para produzir proteínas codificadas por genes exógenos.

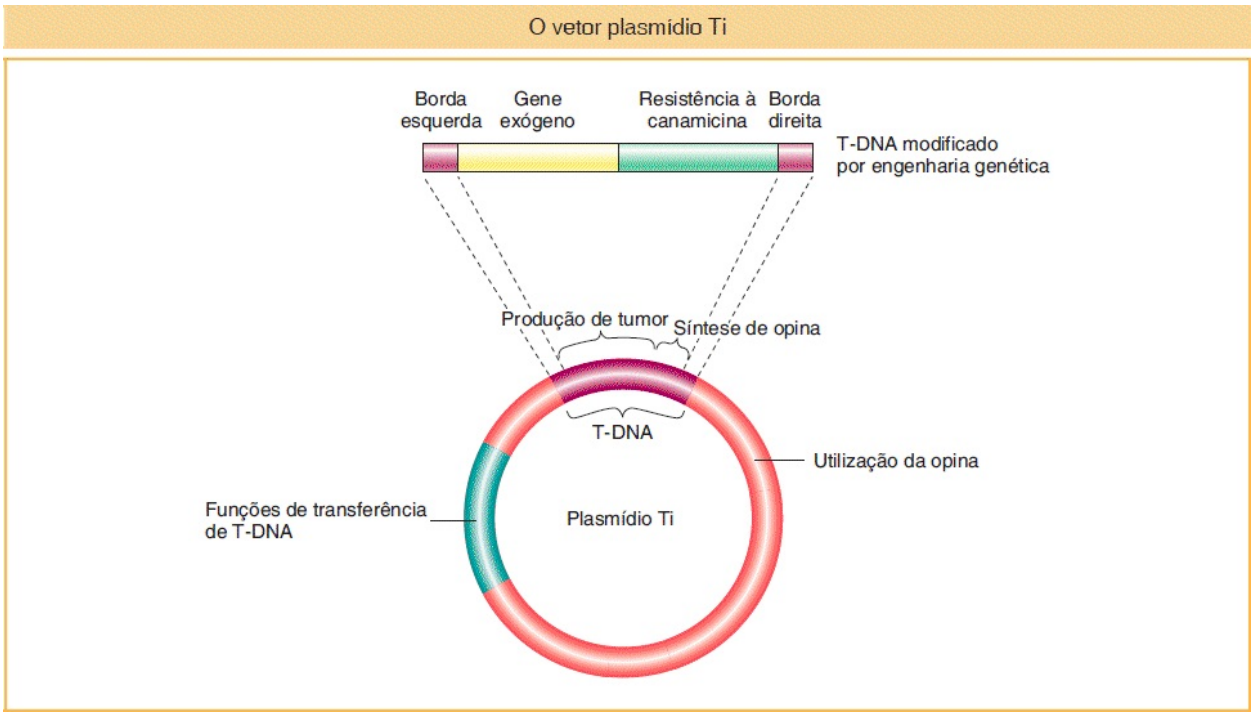


FIGURA 10.24 Representação simplificada das principais regiões do plasmídio Ti de *A. tumefaciens* contendo um T-DNA modificado por engenharia genética.

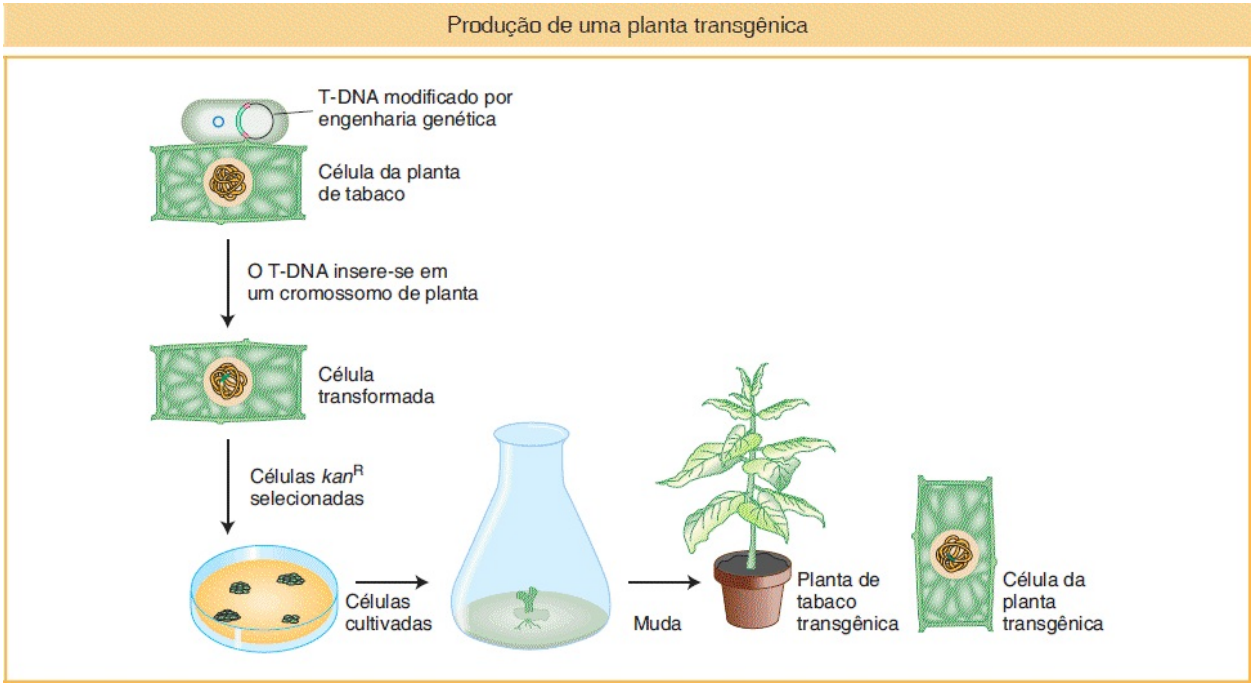


FIGURA 10.25 A inserção de T-DNA nos cromossomos de uma planta. A incubação de discos de folhas com a bactéria *A. tumefaciens* contendo um T-DNA modificado por engenharia genética leva às células das folhas com o T-DNA em seu genoma, que são capazes de crescer em placas de ágar e que podem ser

induzidas a se diferenciarem em plantas do tabaco transgênicas.

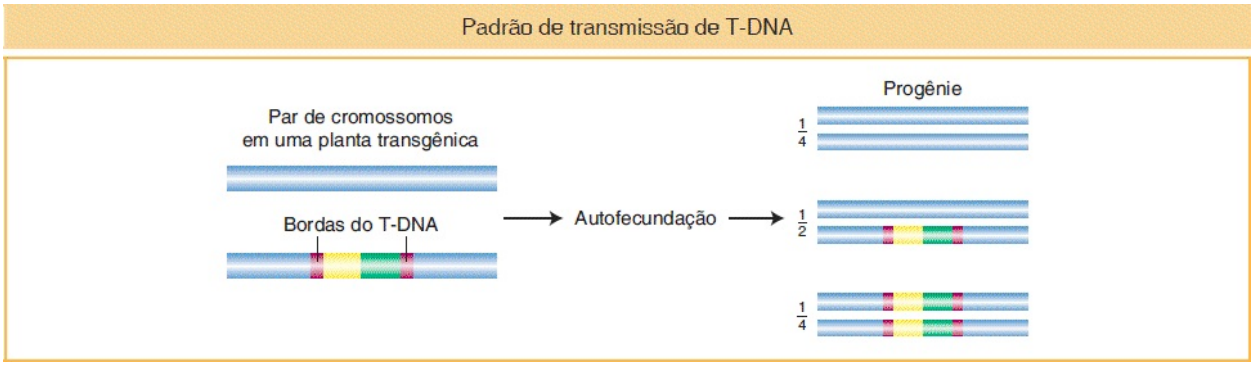


FIGURA 10.26 A região do T-DNA e qualquer DNA inserido no cromossomo de uma planta em uma planta transgênica são transmitidos em um padrão de herança mendeliana.

Engenharia genética em animais

Tecnologias transgênicas estão sendo empregadas atualmente com muitos sistemas de animais-modelo. Enfocaremos os dois modelos de animais bastante utilizados para a pesquisa genética básica: o nematódeo *Caenorhabditis elegans* e o camundongo *Mus musculus*. Um método comumente utilizado para transformar um terceiro organismo-modelo, a mosca-das-frutas *Drosophila melanogaster*, está descrito no Capítulo 15. Versões de muitas das técnicas consideradas até agora também podem ser aplicadas nesses sistemas animais.

Transgênese em *C. elegans*. O método utilizado para introduzir transgenes em *C. elegans* é simples: DNA transgênicos são injetados diretamente no organismo, tipicamente como plasmídios, fosmídios ou outros DNA clonados em bactérias. A estratégia de injeção é determinada pela biologia reprodutiva do verme. As gônadas do verme são sinciciais, o que significa que existem muitos núcleos dentro da mesma célula gonadal. Uma célula sincicial é uma grande proporção de um braço da gônada, e a outra célula sincicial é a maior parte do outro braço (Figura 10.27 A). Esses núcleos não formam células individuais até a meiose, quando iniciam sua transformação em ovócitos ou espermatozoides individuais. Uma solução de DNA é injetada na região sincicial de um dos braços, expondo, assim, mais de 100 núcleos ao DNA transformador. Ao acaso, alguns desses

núcleos incorporarão o DNA (lembre que a membrana nuclear é fragmentada durante a divisão e, assim, o citoplasma dentro do qual o DNA é injetado se torna contínuo com o nucleoplasma). Tipicamente, o DNA transgênico forma *arranjos extracromossômicos* de múltiplas cópias (Figura 10.27 B), que existem como unidades independentes fora dos cromossomos. Mais raramente, os transgenes se tornarão integrados em uma posição ectópica em um cromossomo, ainda como um arranjo de múltiplas cópias. Infelizmente, as sequências podem se misturar aos arranjos, complicando o trabalho do pesquisador.

Transgênese em *M. musculus*. Camundongos são os modelos mais importantes para a genética de mamíferos. O que é mais excitante é que uma grande parte da tecnologia desenvolvida em camundongos é potencialmente aplicável para os seres humanos. Existem duas estratégias para a transgênese em camundongos, cada uma apresentando suas vantagens e desvantagens:

- *Inserções ectópicas.* Os transgenes são inseridos aleatoriamente no genoma, normalmente como arranjos de múltiplas cópias
- *Direcionamento gênico.* A sequência do transgene é inserida em um local ocupado por uma sequência homóloga no genoma. Ou seja, o transgene substitui seu correspondente homólogo normal.

Inserções ectópicas. Para inserir transgenes em locais aleatórios, o procedimento é simplesmente injetar uma solução de DNA clonado em bactéria no núcleo de um zigoto (Figura 10.28 A). Diversos zigotos injetados são inseridos no oviduto de uma fêmea, onde alguns se desenvolverão em filhotes de camundongo. Em algum estágio posterior, o transgene se integrará aos cromossomos de núcleos aleatórios. Ocasionalmente, as células transgênicas formam parte da linhagem germinativa e, nesses casos, um embrião injetado se desenvolverá em um camundongo adulto cujas células germinativas contêm o transgene inserido em alguma posição aleatória em um dos cromossomos (Figura 10.28 B). Parte da progênie desses adultos herdará o transgene em todas as células. Haverá um arranjo de múltiplas cópias do gene em cada ponto de inserção, mas a localização, o tamanho e a estrutura dos arranjos serão diferentes em cada evento

de integração. A técnica dá origem a alguns problemas: (1) o padrão de expressão dos genes aleatoriamente inseridos pode ser anormal (denominado **efeito de posição**), tendo em vista que o ambiente cromossômico local não apresenta as sequências reguladoras normais do gene (ver Capítulo 12 para mais sobre o efeito de posição), e (2) podem ocorrer rearranjos de DNA nos arranjos de múltiplas cópias (em essência, mutando as sequências). Não obstante, essa técnica é muito mais eficiente e menos trabalhosa do que o direcionamento gênico.

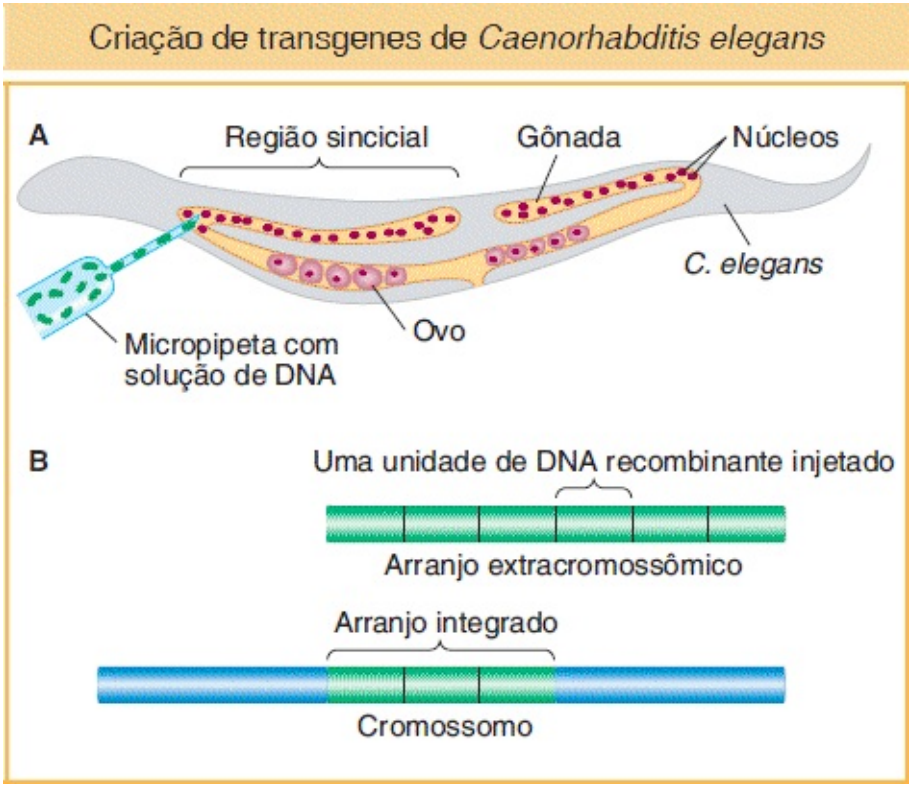


FIGURA 10.27 Transgenes de *C. elegans* são criados por meio da injeção do DNA transgênico diretamente em uma gônada. **A.** Método de injeção. **B.** Os dois tipos principais de resultados transgênicos: arranjos extracromossômicos e arranjos integrados em locais cromossômicos ectópicos.

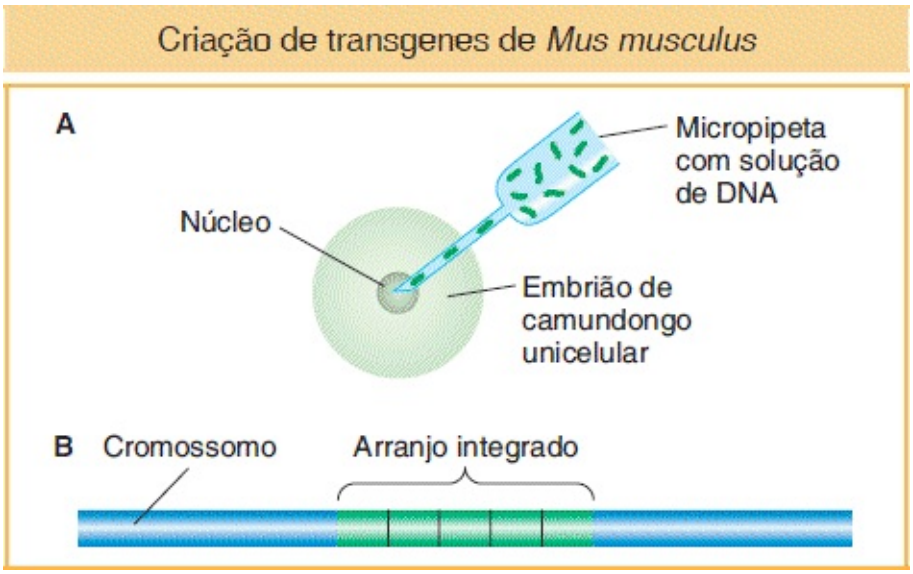


FIGURA 10.28 Transgenes de *M. musculus* são criados por meio da injeção de DNA clonado em ovos fertilizados e da inserção subsequente em locais cromossômicos ectópicos. **A.** Método de injeção. **B.** Um integrante ectópico típico, com múltiplas cópias do transgene recombinante inseridas em um arranjo.

Direcionamento gênico. O direcionamento gênico possibilita que os pesquisadores eliminem um gene ou modifiquem a função que ele codifica. Em uma aplicação, denominada **substituição gênica**, um alelo mutante pode ser reparado por meio de sua substituição por um alelo do tipo selvagem em sua localização cromossômica normal. A substituição gênica evita tanto o efeito de posição quanto os rearranjos de DNA associados à inserção ectópica, tendo em vista que uma única cópia do gene é inserida em seu ambiente cromossômico normal. Contrariamente, um gene pode ser inativado por meio da substituição de um gene normal pelo gene inativo. Uma referida inativação direcionada é denominada **nocaute (*knockout*) gênico**.

O direcionamento gênico em camundongo é realizado em células-tronco embrionárias (células ES) cultivadas. Em geral, uma célula-tronco é uma célula indiferenciada em um determinado tecido ou órgão que se divide assimetricamente para produzir uma progênie de células-tronco e uma célula que irá se diferenciar em um tipo de célula terminal. As células ES são células-tronco especiais que podem se diferenciar para formar qualquer tipo de célula do corpo — incluindo, de maior importância, a linhagem germinativa.

Para ilustrar o processo de direcionamento gênico, observamos como ele alcança um de seus desfechos típicos — a saber, a substituição de um gene normal pelo gene inativo, ou nocaute gênico. O processo requer três estágios:

1. Um gene inativo é direcionado para substituir o gene funcional em uma cultura de células ES, produzindo células ES que contêm um nocaute gênico (Figura 10.29).
2. As células ES que contêm o gene inativo são transferidas para embriões de camundongos (Figura 10.30).
3. Os camundongos transgênicos são identificados e criados para produzir camundongos de genótipo conhecido.

Estágio 1: a versão inativa do gene é preparada por meio da inserção de um segmento do DNA que interrompe cópias do gene clonado. Em seguida construtos de DNA que contêm o gene defeituoso são injetados nos núcleos de células ES cultivadas. O gene defeituoso se insere com muito mais frequência em sítios não homólogos (ectópicos) do que em sítios homólogos (Figura 10.29 B) e, assim, a próxima etapa é selecionar as células raras nas quais o gene defeituoso substituiu o gene funcional, conforme desejado. Como é possível selecionar células ES que contêm um gene raro substituído? O engenheiro genético pode incluir alelos de resistência a fármacos no construto de DNA arranjados de tal modo que as substituições possam ser distinguidas das inserções ectópicas. Um exemplo está demonstrado na Figura 10.29 C.

Estágio 2: as células ES que contêm uma cópia do gene de interesse interrompido (ou seja, o gene nocaute) são injetadas em um embrião em estágio de blastocisto, que em seguida é implantado em uma mãe substituta (Figura 10.30 A). Algumas das células ES podem se tornar incorporadas no embrião hospedeiro e, se isso acontecer, o camundongo que se desenvolve será quimérico — ou seja, conterá células de duas linhagens diferentes de camundongos. Quando o camundongo quimérico alcança a idade adulta, é cruzado com um camundongo normal. Se o camundongo quimérico houver captado as células ES (com o nocaute gênico) nas células da linhagem germinativa, alguns indivíduos da progênie então resultante

herdarão o gene nocaute em todas as suas células. Camundongos da mesma ninhada identificados como heterozigotos em relação à versão do gene nocaute de interesse são cruzados em seguida para produzir camundongos que são homozigotos em relação ao alelo nocaute. (Se o gene for essencial, os homozigotos serão letais e nenhum será obtido a partir desse cruzamento.) (Figura 10.30 B.)

> CONCEITO-CHAVE Foram desenvolvidas técnicas transgênicas em linhagens germinativas para todas as espécies eucarióticas bem-estudadas. Essas técnicas dependem de uma compreensão sobre a biologia reprodutiva da espécie receptora.

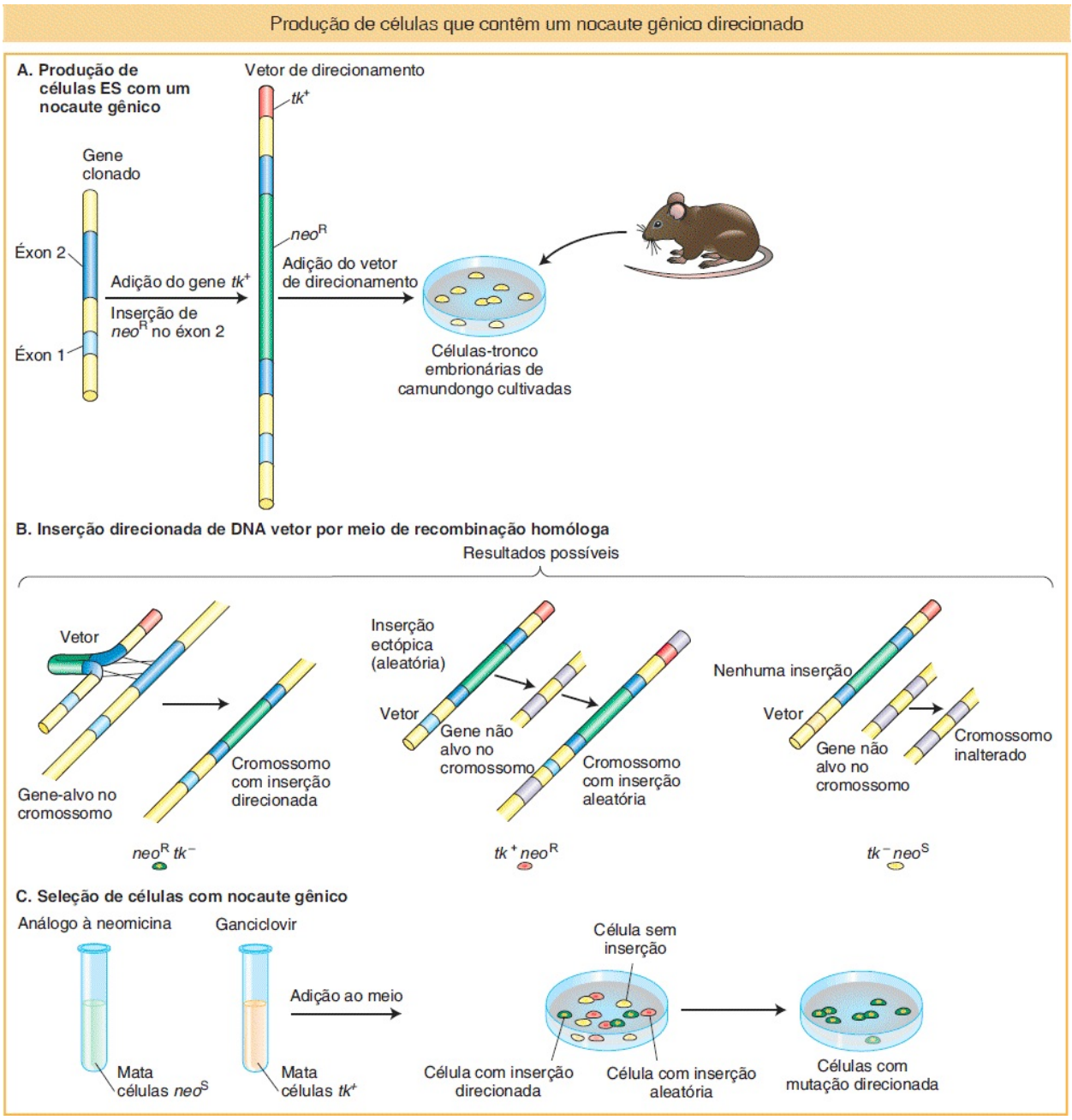


FIGURA 10.29 Produção de células que contêm uma mutação em um gene específico, conhecida como uma mutação direcionada ou nocaute gênico. **A.** Cópias de um gene clonado são alteradas *in vitro* para produzir o vetor que atingirá o alvo. O gene demonstrado aqui foi inativado por meio da inserção do gene de resistência à neomicina (neo^R) em uma região codificadora da proteína (éxon 2) do gene e foi inserido em um vetor. O gene neo^R posteriormente atuará como um marcador para indicar que o DNA vetor se fixou em um cromossomo. O vetor também foi modificado por engenharia genética para carregar um segundo marcador em uma extremidade: o gene *tk* do herpes. Estes marcadores são padrão, mas outros podem ser utilizados. Quando um vetor, com seus dois marcadores, está completo, é introduzido em células isoladas de um embrião de camundongo. **B.** Quando ocorre recombinação homóloga (*esquerda*), as regiões homólogas no vetor,

juntamente com qualquer DNA entre elas, mas excluindo o marcador na extremidade, assumem o lugar do gene original. Este evento é importante em virtude de as sequências do vetor atuarem como um marcador útil para detectar a presença deste gene mutante. Entretanto, em muitas células, o vetor total (completo com o marcador extra na extremidade) se insere ectopicamente (*parte intermediária*) ou não se integra (*direita*). **C.** Para isolar células que carregam uma mutação direcionada, todas as células são colocadas em um meio que contém fármacos selecionados — aqui, um análogo da neomicina (G418) e ganciclovir. G418 é letal para as células, exceto se elas carrearem um gene neo^R funcional, e, assim, elimina as células nas quais não ocorreu a integração do DNA do vetor (*amarelo*). Enquanto isso, o ganciclovir mata quaisquer células que abriguem o gene *tk*, eliminando, assim, as células que contêm um vetor aleatoriamente integrado (*vermelho*). Por consequência, virtualmente as únicas células que sobrevivem e proliferam são aquelas que têm a inserção direcionada (*verde*).

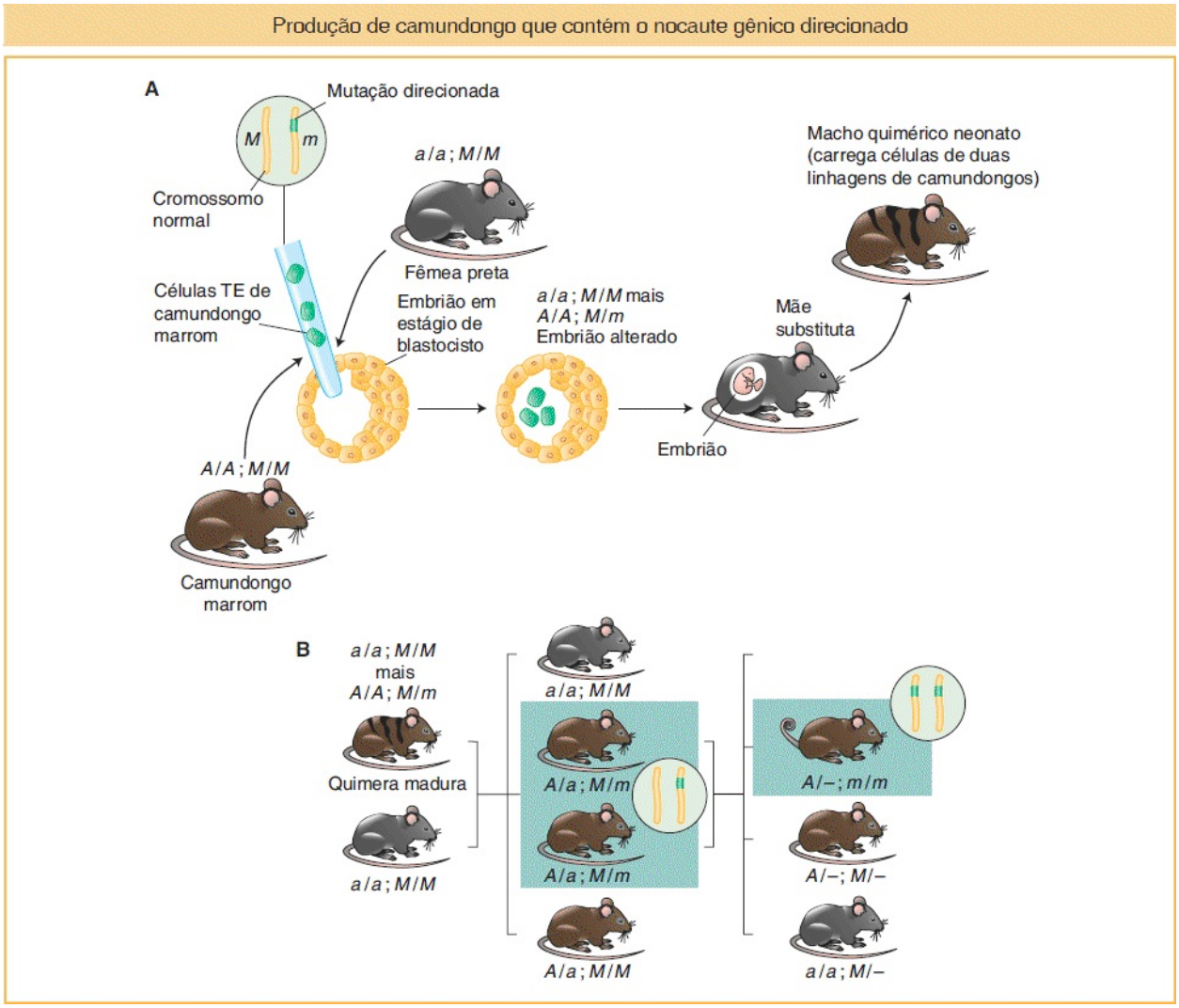


FIGURA 10.30 Um camundongo nocaute é produzido por meio da inserção de células ES que carreiam a mutação direcionada. **A.** As células-tronco embrionárias (ES) são isoladas a partir de uma linhagem de camundongos aguti (*marrom*) (*A/A*) e alteradas para carregar uma mutação direcionada (*m*) em um

cromossomo. Em seguida, as células ES são inseridas em embriões jovens, um dos quais está demonstrado. A cor da pelagem dos futuros neonatos é um guia para saber se as células ES sobreviveram no embrião. Portanto, as células ES são tipicamente colocadas em embriões que, na ausência das células ES, adquiririam uma pelagem totalmente preta. Os referidos embriões são obtidos a partir de uma linhagem preta que não tem o alelo aguti dominante (*a/a*). Os embriões que contêm as células ES crescem a termo em mães substitutas. O matiz aguti entremeado com preto indica aqueles neonatos nos quais as células ES sobreviveram e proliferaram. (Os referidos camundongos são denominados *quimeras*, tendo em vista que contêm células derivadas de duas linhagens diferentes de camundongos.) A pelagem preta sólida, contrariamente, indica que as células ES pereceram, e estes camundongos são excluídos. *A* representa aguti; *a*, preto; *m* é a mutação direcionada; e *M* é o seu alelo do tipo selvagem. **B.** Machos quiméricos são cruzados com fêmeas pretas (não aguti). A progênie é triada à procura de evidências da mutação direcionada (*verde na inserção*) no gene de interesse. O exame direto dos genes em camundongos aguti revela quais daqueles animais (*no quadro*) herdaram a mutação direcionada. Machos e fêmeas que carregam a mutação são cruzados entre si para produzir camundongos cujas células carregam a mutação escolhida em ambas as cópias do gene-alvo (*inserção*) e, portanto, apresentam ausência de um gene funcional. Os referidos animais (*no quadro*) são definitivamente identificados por meio de análises diretas de seu DNA. O nocaute neste caso resulta em um fenótipo com cauda enrolada.

RESUMO

O DNA recombinante é construído em laboratório para possibilitar que pesquisadores amplifiquem e analisem segmentos de DNA (DNA doador) de qualquer genoma ou de cópias de DNA dos mRNA. Três fontes de DNA doador são (1) o genoma inteiro digerido com uma enzima de restrição, (2) os produtos da PCR de regiões específicas do DNA definidas pelas sequências de *primer* flanqueadores, e (3) as cópias de cDNA de mRNA.

A reação da cadeia de polimerase é um método poderoso para a amplificação direta de uma sequência relativamente pequena de DNA a partir de uma mistura complexa de DNA, sem a necessidade de uma célula hospedeira ou de muito material inicial. A chave é possuir *primers* que são complementares às regiões flanqueadoras em cada um dos dois filamentos de DNA. Essas regiões atuam como sítios para a polimerização. Múltiplas rodadas de desnaturação, pareamento de *primers* e polimerização amplificam exponencialmente a sequência de interesse.

Para inserir o DNA doador nos vetores, o DNA doador e o DNA vetor são cortados pela mesma endonuclease de restrição em sequências específicas. O

DNA vetor e o doador são unidos por meio do pareamento das extremidades coesivas que resultam da digestão, seguida por ligação para unir covalentemente as moléculas. As moléculas de PCR e cDNA são inseridas em vetores primeiramente por meio da adição de sequências de reconhecimento de endonucleases de restrição à extremidade 5′ dos *primers* da PCR ou por meio da ligação de adaptadores curtos que contêm sítios de restrição em suas extremidades antes da inserção no vetor.

Existe uma ampla variedade de vetores bacterianos. A escolha do vetor depende amplamente do tamanho do fragmento de DNA a ser clonado. Plasmídios são utilizados para clonar pequenos fragmentos de restrição, moléculas de PCR, ou moléculas de cDNA. Fragmentos de tamanho intermediário, tais como aqueles que resultam da digestão de DNA genômico, podem ser clonados em versões modificadas do bacteriófago λ (para insertos de 10 a 15 kb) ou em híbridos de fago e plasmídio denominados fosmídios (para insertos de 35 a 45 kb). Finalmente, cromossomos artificiais bacterianos (BAC) são utilizados de modo rotineiro para clonar fragmentos genômicos muito grandes (aproximadamente 100 a 120 kb).

O construto de DNA vetor-doador é amplificado dentro de células hospedeiras bacterianas como moléculas extracromossômicas que são replicadas quando o hospedeiro está replicando o seu genoma. O resultado da amplificação de plasmídios, fagos e BAC é de clones que contêm múltiplas cópias de cada constructo de DNA recombinante. Contrariamente, apenas um único fosmídio está presente em cada célula bacteriana.

Com frequência, encontrar um clone específico com um gene de interesse requer a triagem de toda uma biblioteca genômica. Uma biblioteca genômica é um conjunto de clones, ligados no mesmo vetor, que em conjunto representam todas as regiões do genoma do organismo em questão. O número de clones que constitui uma biblioteca genômica depende (1) do tamanho do genoma em questão e (2) do tamanho do inserto tolerado pelo sistema de vetor de clonagem em particular. De modo semelhante, uma biblioteca de cDNA é uma representação do conjunto total de mRNA produzido por um determinado tecido ou estágio de desenvolvimento em um determinado organismo.

Sondas de DNA unifilamentar ou RNA marcadas são importantes para identificar sequências semelhantes ou idênticas em misturas complexas de moléculas, seja em bibliotecas genômicas ou de cDNA, ou no *Southern blotting* (DNA) e no *Northern blotting* (RNA). O princípio geral da técnica para a identificação de clones ou fragmentos em gel é criar uma "imagem" das colônias ou das placas em uma cultura em placa de Petri com ágar ou dos ácidos nucleicos que foram separados em um campo elétrico passado por uma matriz de gel. Em seguida o DNA ou RNA é desnaturado e misturado com uma sonda desnaturada que foi marcada com um corante fluorescente ou um marcador radioativo. Após a sonda não ligada ter sido eliminada, a localização da sonda é detectada por meio da observação de sua fluorescência, ou, se radioativa, por meio da exposição da amostra a um filme radiográfico. As localizações da sonda correspondem às localizações do DNA ou RNA relevante na placa de Petri original ou na eletroforese em gel. Os anticorpos marcados são sondas importantes para identificar proteínas específicas em misturas complexas produzidas por bibliotecas de expressão (com insertos de cDNA) ou no *Western blotting*.

Os vastos recursos genômicos estão tornando cada vez mais possível o isolamento de genes unicamente a partir do conhecimento sobre a sua posição em um mapa genético. Dois procedimentos gerais são as estratégias genéticas diretas denominadas clonagem posicional e mapeamento de estrutura fina. Com o sequenciamento do genoma humano e a disponibilidade de famílias com distúrbios hereditários, estratégias de mapeamento de estrutura fina levaram ao isolamento de genes que, quando mutados, produzem a doença humana.

Os transgenes são moléculas de DNA modificadas por engenharia genética que são introduzidas e expressas em células eucarióticas. Eles podem ser utilizados para modificar uma mutação nova ou para estudar as sequências reguladoras que constituem parte de um gene. Os transgenes podem ser introduzidos como moléculas extracromossômicas, ou podem ser integrados em um cromossomo, seja em locais aleatórios (ectópicos), ou no lugar do gene homólogo, dependendo do sistema. Tipicamente, os mecanismos utilizados para introduzir um transgene dependem da compreensão e da exploração da biologia reprodutiva do organismo.

TERMOS-CHAVE

amplificação
anticorpo
autorradiograma
biblioteca de cDNA
biblioteca genômica
caminhada cromossômica (*chromosome walk*)
clonagem de DNA
clonagem posicional
complementação funcional (resgate de mutante)
cromossomo artificial bacteriano (BAC)
DNA complementar (cDNA)
DNA doador
DNA ligase
DNA recombinante
ectopicamente
efeito de posição
eletroforese em gel
engenharia genética
enzima de restrição
fosmídio
fragmento de restrição
genômica
hibridização
mapeamento fino
nocaute (*knockout*) gênico
Northern blotting
organismo geneticamente modificado (OGM)
organismo transgênico
palíndromo de DNA
plasmídio Ti

reação da cadeia de polimerase (PCR)
resgate de mutante (complementação funcional)
RNA *blotting*
sequenciamento de Sanger (didesóxi)
sequenciamento didesóxi (de Sanger)
sonda
Southern blotting
substituição gênica
tecnologia do DNA
transgene
vetores

PROBLEMAS RESOLVIDOS

Problema resolvido 1. No Capítulo 9, estudamos a estrutura das moléculas de tRNA. Suponha que você deseja clonar gene de fungo que codifica um determinado tRNA. Você possui uma amostra de tRNA purificado e um plasmídio de *E. coli* que contém um único sítio de corte de *Eco*RI em um gene tet^R (resistência à tetraciclina), bem como um gene de resistência à ampicilina (amp^R). Como você pode clonar o gene de interesse?

Solução

Você pode utilizar o próprio tRNA ou uma cópia do cDNA clonado dele para investigar o DNA que contém o gene. Um método é digerir o DNA genômico com *Eco*RI e em seguida misturá-lo com o plasmídio, que você também deve cortar com a *Eco*RI. Após a transformação de um receptor amp^S tet^S, selecione colônias de Amp^R, que indicam a transformação de sucesso. Dessas colônias de Amp^R, selecione as colônias que são Tet^S. Essas colônias Tet^S conterão vetores com insertos no gene tet^R, e é necessário um grande número deles para produzir a biblioteca. Teste a biblioteca por meio da utilização do tRNA como sonda. Aqueles clones que hibridizam com a sonda conterão o gene de interesse.

Alternativamente, você pode submeter o DNA genômico digerido por *Eco*RI à

eletroforese em gel e em seguida identificar a banda correta hibridizando a sonda com o tRNA. Essa região do gel pode ser cortada e utilizada como uma fonte de DNA enriquecido para clonar no plasmídio cortado com *Eco*RI. Em seguida você hibridiza esses clones com o tRNA para confirmar que esses clones contêm o gene de interesse.

Problema resolvido 2. Você isolou um gene de levedura que atua na síntese do aminoácido leucina, e você formula a hipótese de que ele apresenta a mesma função de um gene correlato de *E. coli*. Como você utilizaria a complementação funcional/resgate de mutante para testar a sua teoria?

Solução

Primeiramente, assumiremos que o gene de interesse da levedura está em um fragmento digerido por *Eco*RI. Você utilizará técnicas de DNA recombinante para inserir esse fragmento no sítio poliligador de um plasmídio bacteriano que foi cortado com *Eco*RI e tratado com ligase para formar novamente o plasmídio circular. Esse plasmídio também deve conter um marcador selecionável para resistência a antibióticos, tal como amp^R (ver Figura 10.9). Em seguida, você precisa transformar esse plasmídio recombinante em mutantes de *E. coli leu*$^-$. Entretanto, tendo em vista que existem quatro genes necessários para a biossíntese da leucina e você não sabe qual pode ser o seu gene de levedura, você precisa testar para complementação em quatro linhagens de *E. coli* mutantes, cada uma com uma mutação em um gene diferente dos quatro genes (denominados *leuA*, *leuB*, *leuC*, e *leuD*). Cultive as quatro linhagens de *E. coli* em separado, e realize um experimento de transformação no qual o mesmo plasmídio recombinante é introduzido em cada uma das quatro linhagens. Em seguida plaqueie os transformantes em placas com ágar que contenham o antibiótico ampicilina (assim, apenas as células que captaram o plasmídio conseguem crescer), mas que não contenham leucina. Se você observar colônias crescendo em uma das quatro placas que contêm células de *E. coli* transformadas, não apenas você pode concluir que isolou um gene de levedura envolvido na biossíntese da leucina, como também pode determinar qual é o gene leu (*leuA*, *leuB*, *leuC*, ou *leuD*). Ou

seja, se o gene da levedura for *leuA*, ele somente complementará o mutante de *E. coli leuA*$^{-}$.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. A Figura 10.1 demonstra que fragmentos de DNA específicos podem ser sintetizados *in vitro* antes da clonagem. Quais são dois modos de sintetizar insertos de DNA para DNA recombinante *in vitro*?

2. Na Figura 10.4, por que o cDNA é produzido apenas a partir do mRNA e não também a partir dos tRNA e RNA ribossômicos?

3. Redesenhe a Figura 10.6 com o objetivo de adicionar uma extremidade *Eco*RI e uma extremidade *Xho*I. A seguir apresentamos a sequência de reconhecimento de *Xho*I.

Sequência de reconhecimento:
...CTCGAG... ...GAGCTC...

Após o corte:
...C TCGAG... ...GAGCT C...

4. Redesenhe a Figura 10.7 de modo que o cDNA possa ser inserido em um sítio *Xho*I de um vetor, em vez de em um sítio *Eco*RI, conforme demonstrado.

5. Na Figura 10.10, determine aproximadamente quantos clones BAC são necessários para proporcionar 1× a cobertura:
a. do genoma de levedura (12 Mpb).
b. do genoma de *E. coli* (4,1 Mpb).
c. do genoma da mosca-das-frutas (130 Mpb).

6. Na Figura 10.14, por que o DNA migra para o anodo (polo +)?

7. Na Figura 10.17 A, por que os fragmentos de DNA são de diferentes comprimentos e todos terminam em um resíduo A sintetizado?

8. Conforme você verá no Capítulo 15, a maior parte dos genomas de eucariotos superiores (plantas e animais) é preenchida por sequências de DNA que estão presentes em centenas, até mesmo milhares de cópias por todos os cromossomos. No procedimento de caminhada cromossômica demonstrado na Figura 10.19, como o experimentador saberia se o fragmento que está utilizando para "caminhar" até o próximo BAC ou fago é repetitivo? O DNA repetitivo pode ser utilizado em uma caminhada cromossômica?

9. Redesenhe a Figura 10.23 para incluir as posições dos *crossovers* únicos e duplos.

10. Na Figura 10.25, por que apenas as células de plantas que apresentam insertos de T-DNA em seus cromossomos crescem em placas com ágar? Todas as células de uma planta transgênica cultivada a partir de um aglomerado de células contêm o T-DNA? Justifique a sua resposta.

11. Na Figura 10.27, qual é a diferença entre DNA extracromossômico e os arranjos integrados de DNA? Os últimos são ectópicos? Qual é a distinção da região sincicial que a torna um bom local para injetar DNA?

PROBLEMAS BÁSICOS

12. A partir deste capítulo, faça uma lista de todos os exemplos **(a)** de hibridização de DNA unifilamentares e **(b)** de proteínas que se ligam ao DNA e em seguida atuam sobre ele.

13. Compare e contraste a utilização da palavra *recombinante* conforme utilizada nas frases **(a)** "DNA recombinante" e **(b)** "frequência de recombinantes".

14. Por que a ligase é necessária para produzir DNA recombinante? Qual seria a consequência imediata no processo de clonagem se uma pessoa se esquecesse de adicioná-la?

15. No processo de PCR, se presumirmos que cada ciclo demora 5 minutos, quantas vezes a amplificação seria realizada em 1 hora?

16. A posição do gene para a proteína actina no fungo haploide *Neurospora* é conhecida a partir da sequência genômica completa. Se você tivesse um mutante de crescimento lento que você suspeitasse ser um mutante da actina e desejasse verificar se era um mutante da actina, você **(a)** clonaria o mutante por meio da utilização de sítios de restrição convenientes que flanqueiam o gene da actina e em seguida realizaria o seu sequenciamento, ou **(b)** amplificaria o gene mutante por meio da utilização de PCR e em seguida realizaria o seu sequenciamento?

17. Você obtém a sequência de DNA de um mutante de um gene de 2 kb no qual você está interessado e ela demonstra diferenças de bases em três posições, todas em diferentes códons. Uma é uma alteração silenciosa, mas as outras duas são alterações de sentido trocado (elas codificam novos aminoácidos). Como você demonstraria que essas alterações são mutações reais e não erros de sequenciamento? (Presuma que o sequenciamento é aproximadamente 99,9% preciso.)

18. Em uma transformação de T-DNA de uma planta com um transgene de um fungo (não observado em plantas), a planta presumidamente transgênica não expressa o fenótipo do transgene esperado. Como você demonstraria que o transgene de fato está presente? Como você demonstraria que o transgene foi expresso?

19. Como você produziria um camundongo homozigoto para um transgene de hormônio de crescimento de ratos?

20. Por que o cDNA, e não o DNA genômico, foi utilizado na clonagem comercial do gene da insulina humana?

21. Após o DNA de *Drosophila* ter sido tratado com uma enzima de restrição, os fragmentos são inseridos em plasmídios e selecionados como clones em *E. coli*. Com a utilização dessa técnica de *shotgun*, cada sequência do DNA de *Drosophila* em uma biblioteca pode ser recuperada.

a. Como você identificaria um clone que contém DNA que codifica a proteína actina, cuja sequência de aminoácidos é conhecida?

b. Como você identificaria um clone que codifica um tRNA específico?

22. Em qualquer célula eucariótica transformada (digamos, de *Saccharomyces cerevisiae*), como você poderia dizer se o DNA transformador (carregado em um vetor bacteriano circular):

a. substituiu o gene residente do receptor por meio de *crossing over* duplo ou de *crossing over* único?

b. foi inserido ectopicamente?

23. Em um gel de eletroforese ao longo do qual é aplicado um campo elétrico pulsado alternado poderoso, o DNA do fungo haploide *Neurospora crassa* (n = 7) se movimenta lentamente, mas ao final forma sete bandas, que representam frações de DNA que são de diferentes tamanhos e que, portanto, movimentaram-se em diferentes velocidades. Essas bandas presumidamente são os sete cromossomos. Como você demonstraria qual banda corresponde a qual cromossomo?

24. A proteína codificada pelo gene da fibrose cística apresenta 1.480 aminoácidos de comprimento, ainda que o gene abranja 250 kb. Como essa diferença é possível?

25. Em uma levedura, você sequenciou um segmento de DNA do tipo selvagem e ele claramente contém um gene, mas você não sabe que gene é. Portanto, para investigar adicionalmente, você gostaria de encontrar seu fenótipo mutante. Como você utilizaria o gene do tipo selvagem clonado para tanto? Demonstre claramente as suas etapas experimentais.

26. Por que é necessário utilizar uma DNA polimerase especial (*Taq* polimerase) na PCR?

27. Em relação a cada um dos objetivos experimentais a seguir, é preferível a PCR ou a clonagem do gene e por quê?

a. Isolar o mesmo gene de 20 indivíduos.

b. Isolar 100 genes do mesmo indivíduo.

c. Isolar um gene de camundongo quando você possui um fragmento de gene de rato.

28. No *Northern blotting*, a eletroforese é utilizada para analisar quais

moléculas biológicas? Qual tipo de sonda é utilizado para identificar a(s) molécula(s)-alvo?

29. Uma característica que virtualmente todos os vetores plasmidiais apresentam em comum é o poliligador (também denominado sítio de clonagem múltipla). Explique o que é um poliligador e por que essa é uma característica tão importante.

30. Uma segunda característica que virtualmente todos os vetores plasmidiais apresentam em comum é o marcador selecionável. Explique o que é isso e por que ele é uma característica tão importante.

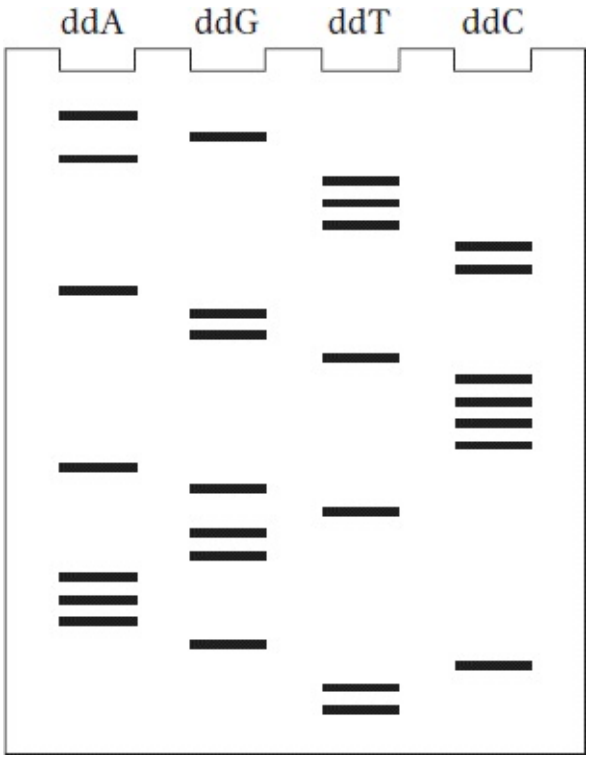
PROBLEMAS DESAFIADORES

31. A prototrofia com frequência é o fenótipo selecionado para detectar transformantes. Células prototróficas são utilizadas para a extração de DNA doador; em seguida, esse DNA é clonado e os clones são adicionados a uma cultura receptora auxotrófica. Transformantes bem-sucedidos são identificados por meio do plaqueamento da cultura receptora em meio mínimo e da observação das colônias. Qual desenho experimental você utilizaria para assegurar que uma colônia que você espera ser um transformante não seja, de fato:

a. uma célula prototrófica que entrou na cultura receptora como um contaminante?

b. um revertente (mutação de retorno à prototrofia por meio de uma segunda mutação no gene originalmente mutado) da mutação auxotrófica?

32. Um fragmento de DNA clonado foi sequenciado por meio da utilização do método de término de cadeia didesóxi. Uma parte do autorradiograma do gel de sequenciamento está representada aqui.



a. Deduza a sequência de nucleotídios da cadeia de nucleotídios do DNA sintetizada a partir do *primer*. Rotule as extremidades 5′ e 3′.

b. Deduza a sequência de nucleotídios da cadeia de nucleotídios do DNA utilizada como o filamento-molde. Rotule as extremidades 5′ e 3′.

c. Escreva a sequência de nucleotídios da dupla-hélice de DNA (rotule as extremidades 5′ e 3′).

33. O clone de cDNA em relação ao gene humano que codifica a tirosinase foi marcado radioativamente e utilizado em uma análise de *Southern* de DNA genômico digerido por *Eco*RI de camundongos do tipo selvagem. Observou-se que três fragmentos de camundongo eram radioativos (estavam ligados à sonda). Quando camundongos albinos foram utilizados nessa análise de *Southern*, nenhum fragmento genômico ligou-se à sonda. Explique esses resultados em relação à natureza dos alelos de camundongo do tipo selvagem e mutante.

34. Foram obtidas plantas de tabaco transgênicas nas quais o vetor plasmídio Ti foi desenhado para inserir o gene de interesse mais um gene adjacente de resistência à canamicina. A herança da inserção cromossômica foi seguida

pelo teste da progênie em relação à resistência à canamicina. Duas plantas tipificaram os resultados obtidos em geral. Quando a planta 1 foi submetida ao retrocruzamento com tabaco do tipo selvagem, 50% da progênie foi resistente à canamicina e 50% foi sensível. Quando a planta 2 foi submetida ao retrocruzamento com o tipo selvagem, 75% da progênie foi resistente à canamicina e 25% foi sensível. Qual deve ter sido a diferença entre as duas plantas transgênicas? O que você preveria sobre a situação a respeito do gene de interesse?

35. Uma mutação para fibrose cística em um determinado heredograma ocorre em virtude de uma alteração em um único par de nucleotídios. Essa alteração destrói um sítio de restrição de *Eco*RI normalmente observado nessa posição. Como você utilizaria essas informações no aconselhamento de membros dessa família a respeito da sua probabilidade de serem portadores? Declare os experimentos precisos necessários. Presuma que você encontra uma mulher nessa família que é portadora, e ocorre que ela é casada com um homem não relacionado que também é um heterozigoto em relação à fibrose cística, mas, no caso dele, é uma mutação diferente no mesmo gene. Como você aconselharia esse casal a respeito dos riscos de um filho apresentar a fibrose cística?

36. A glicuronidase bacteriana converte uma substância incolor denominada X-Gluc em um pigmento azul índigo brilhante. O gene para a glicuronidase também atua em plantas se for fornecida uma região promotora à planta. Como você utilizaria esse gene como um gene repórter para encontrar os tecidos nos quais um gene de planta que você acabou de clonar normalmente é ativo? (Presuma que o X-Gluc é facilmente captado pelos tecidos da planta.)

37. A planta *Arabidopsis thaliana* foi transformada por meio da utilização do plasmídio Ti, dentro do qual havia sido inserido um gene de resistência à canamicina na região T-DNA. Duas colônias resistentes à canamicina (A e B) foram selecionadas, e as plantas foram regeneradas a partir delas. Foi possibilitada a autopolinização das plantas, e os resultados foram como

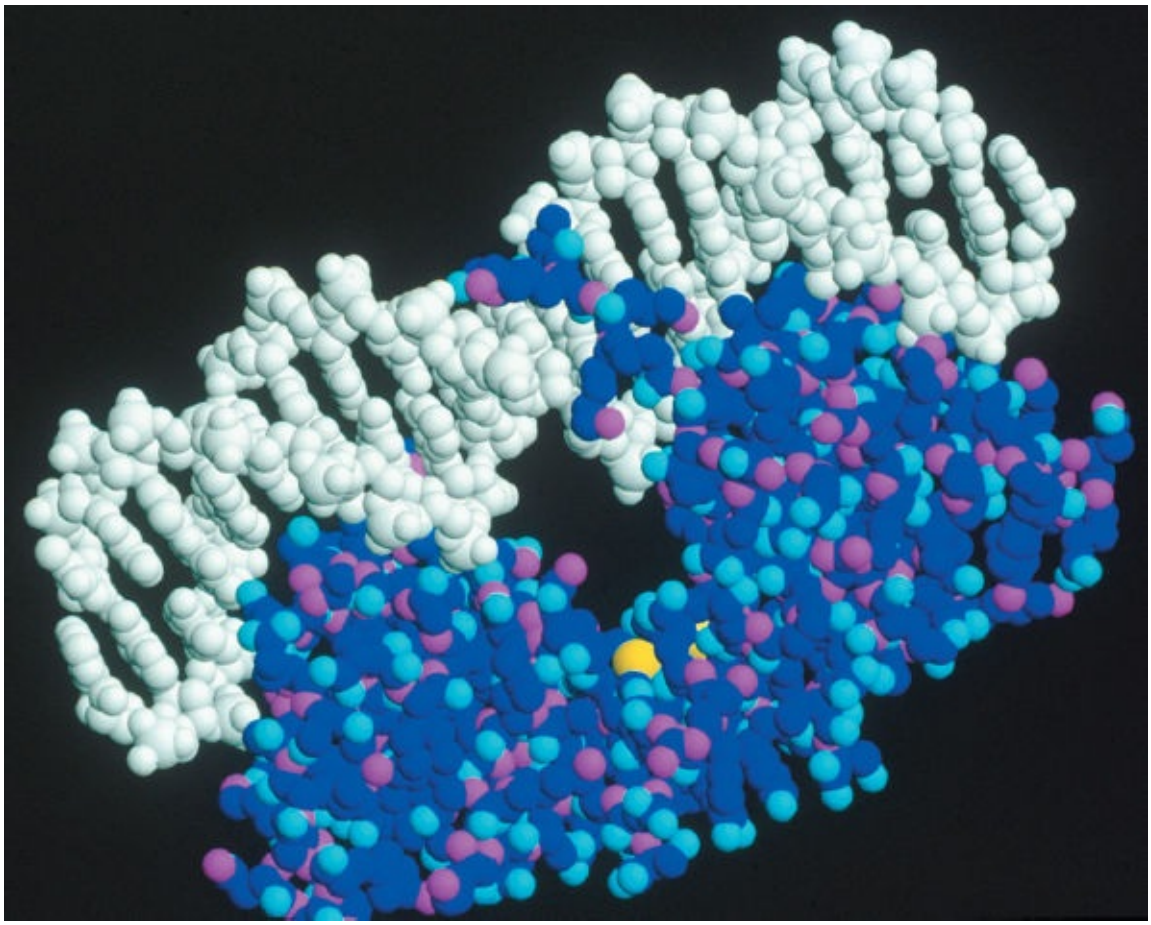
segue:

Planta A autopolinizada	→	$\frac{3}{4}$ da progênie resistente à canamicina
		$\frac{1}{4}$ da progênie sensível à canamicina
Planta B autopolinizada	→	$\frac{15}{16}$ da progênie resistente à canamicina
		$\frac{1}{16}$ da progênie resistente à canamicina

a. Desenhe os cromossomos da planta relevantes em ambas as plantas.
b. Explique as duas proporções diferentes.

CAPÍTULO 11

Regulação da Expressão Gênica em Bactérias e seus Vírus



O controle da expressão gênica é regulado primariamente por proteínas de ligação ao DNA que reconhecem sequências de controle específicas dos genes. Aqui, está modelada a ligação da proteína repressora Lac ao DNA do operador lac. (Kenneth Eward/Science Source.)

TÓPICOS

11.1 Regulação gênica

11.2 Descoberta do sistema lac | Controle negativo

11.3 Repressão catabólica do óperon lac | Controle positivo
11.4 Duplo controle positivo e negativo | Óperon da arabinose
11.5 Vias metabólicas e níveis adicionais de regulação | Atenuação
11.6 Ciclos de vida de bacteriófagos | Mais reguladores, óperons complexos
11.7 Fatores sigma alternativos regulam grandes conjuntos de genes

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Contrastar regulações positiva e negativa da expressão gênica e explicar como ambos os mecanismos controlam a atividade do operon *lac*
- Identificar os componentes de ação trans e de ação cis dos óperons, e prever o efeito de mutações nesses componentes sobre a expressão gênica
- Comparar como moléculas simples acionam alterações na expressão gênica de diferentes óperons em bactérias
- Explicar os papéis das proteínas de ligação ao DNA sequência-específica e as sequências reguladoras do DNA na coordenação da expressão de conjuntos de genes em bactérias e bacteriófagos.

Em dezembro de 1965, o rei da Suécia concedeu o Prêmio Nobel de Fisiologia ou Medicina a François Jacob, Jacques Monod e André Lwoff, do Instituto Pasteur, por suas descobertas sobre como a expressão gênica é regulada (Figura 11.1). O prêmio foi o fruto de uma excepcional colaboração entre três cientistas esplêndidos. Foi também um triunfo sobre grandes obstáculos. Eram poucas as chances de que cada um desses três homens vivesse para ver aquele dia, ainda mais para receber as referidas honrarias.

Vinte e cinco anos antes, Monod era um doutorando na Sorbonne, em Paris, e trabalhava com um fenômeno em bactérias denominado “adaptação enzimática”, algo tão obscuro para alguns que o diretor do laboratório de zoologia onde ele trabalhava declarou: “O que Jacques Monod está realizando é de absolutamente nenhum interesse para a Sorbonne.” Jacob era um estudante de medicina de 19 anos de idade que pretendia tornar-se cirurgião. Lwoff naquela época era um

membro bem-estabelecido do Instituto Pasteur em Paris, chefe do departamento de fisiologia microbiana.

Então, veio a Segunda Guerra Mundial.

Na medida em que a França foi invadida e rapidamente derrotada, Jacob correu até a costa para se juntar às forças francesas livres que se reuniam na Inglaterra. Ele serviu como médico no norte da África e na Normandia até que foi gravemente ferido. Monod se juntou à Resistência Francesa enquanto continuava seu trabalho. Após uma incursão da Gestapo em seu laboratório na Sorbonne, Monod decidiu que trabalhar ali era muito perigoso (seu predecessor na Resistência foi preso e executado), e André Lwoff ofereceu a ele um espaço no Pasteur. Monod, por sua vez, conectou Lwoff com a Resistência.

FIGURA 11.1 François Jacob, Jacques Monod e André Lwoff receberam o Prêmio Nobel de Fisiologia ou Medicina de 1965 por seu trabalho pioneiro sobre como a expressão gênica é regulada. (Instituto Pasteur.)

Após a libertação de Paris, Monod serviu no Exército francês e teve acesso a um artigo de Oswald Avery e colegas que demonstrava que o DNA é o material hereditário em bactérias (ver Capítulo 7). Seu interesse por genética foi reacendido, e ele reencontrou Lwoff depois da guerra. Enquanto isso, as lesões de Jacob eram muito graves para que ele seguisse uma carreira em cirurgia. Inspirado pelo enorme impacto dos antibióticos introduzidos tardiamente durante a guerra, Jacob finalmente decidiu seguir a pesquisa científica. Jacob se aproximou de Lwoff diversas vezes pretendendo uma posição em seu laboratório, mas foi recusado. Ele fez uma última tentativa e encontrou Lwoff em um humor jovial. O cientista sênior disse a Jacob: "Você sabe, acabamos de descobrir a indução do prófago. Você estaria interessado em trabalhar com o fago?" Jacob não tinha ideia sobre o que Lwoff estava falando. Ele gaguejou: "Isso é exatamente o que eu gostaria de fazer."

O elenco estava estabelecido. O desdobramento na década subsequente foi uma das colaborações mais criativas e produtivas na história da genética, cujas descobertas ainda reverberam por toda a biologia atualmente.

Uma das percepções mais importantes ocorreu não no laboratório, mas em um cinema. Lutando para preparar uma palestra, Jacob optou por levar sua esposa, Lise, a uma matinê de domingo. Aborrecido e sonhando acordado, Jacob elaborou uma conexão entre o trabalho que ele estava realizando sobre a indução do prófago e aquele de Monod sobre a indução da síntese enzimática. Jacob foi "envolvido por uma súbita animação misturada a um vago prazer (…) Ambos os experimentos (…) sobre o fago (…) e aquele realizado com Pardee e Monod no sistema da lactose (…) são os mesmos! Mesma situação. Mesmo resultado (…) Em ambos os casos, o gene regula a formação (…) de um repressor que bloqueia a expressão de outros genes e, assim, evita a síntese de galactosidase ou a multiplicação do vírus (…) Onde o repressor pode atuar para interromper tudo de uma vez? A única resposta simples (…) é no próprio DNA!"[1]

E, assim, nascia o conceito de um repressor que atua no DNA para reprimir a indução de genes. Seriam necessários muitos anos antes que os repressores hipotéticos fossem isolados e caracterizados bioquimicamente. Os conceitos trabalhados por Jacob e Monod e explicados neste capítulo — RNA mensageiro,

promotores, operadores, genes reguladores, óperons e proteínas alostéricas — foram deduzidos inteiramente a partir de evidências genéticas, e esses conceitos modelaram o futuro campo da genética molecular.

Walter Gilbert, que isolou o primeiro repressor e que posteriormente recebeu um Prêmio Nobel em Química por coinventar um método de sequenciamento do DNA, explicou o efeito do trabalho de Jacob e Monod naquela época: "A maior parte das descobertas cruciais na ciência são de uma tal natureza tão simples que é muito difícil até mesmo de conceber sem de fato realizar a experiência envolvida na descoberta… A sugestão de Jacob e Monod tornou as coisas que eram completamente obscuras muito simples."[2]

Os conceitos que Jacob e Monod iluminaram muito além das enzimas e dos vírus bacterianos. Eles compreenderam, e foram capazes de articular com eloquência excepcional, como as suas descobertas a respeito da regulação gênica estavam relacionadas aos mistérios gerais da diferenciação celular e do desenvolvimento embrionário em animais. Os dois homens uma vez gracejaram: "qualquer coisa que se observe ser verdadeira para a *E. coli* também deve ser verdadeira para os elefantes."[3] Nos próximos três capítulos, veremos até que grau essa afirmação é verdadeira. Iniciaremos neste capítulo com exemplos bacterianos que ilustram temas-chave e mecanismos-chave na regulação da expressão gênica. Enfocaremos amplamente proteínas reguladoras únicas e os "interruptores" genéticos sobre as quais elas atuam. Em seguida, no Capítulo 12, abordaremos a regulação gênica em células eucarióticas, que envolve um maquinário bioquímico e genético mais complexo. E, finalmente, no Capítulo 13, examinaremos o papel da regulação gênica no desenvolvimento de animais multicelulares. Ali veremos como conjuntos de proteínas reguladoras atuam sobre arranjos de interruptores genéticos para controlar a expressão gênica no tempo e no espaço e coreografar a construção de corpos e partes corporais.

11.1 Regulação gênica

Apesar da sua simplicidade de forma, as bactérias apresentam em comum com os

organismos maiores e mais complexos a necessidade de regular a expressão de seus genes. Um dos principais motivos é que elas são oportunistas nutricionais. Considere como as bactérias obtêm muitos compostos importantes, tais como açúcares, aminoácidos e nucleotídios, necessários para o metabolismo. As bactérias nadam em um mar de possíveis nutrientes. Elas podem adquirir os compostos de que necessitam a partir do ambiente ou sintetizá-los por meio de vias enzimáticas. Mas a síntese desses compostos também requer gasto energético e recursos celulares para produzir as enzimas necessárias para essas vias. Portanto, quando lhes for proporcionada uma opção, as bactérias captarão compostos do ambiente. A seleção natural favorece a eficiência contra o gasto de recursos e de energia. Para serem econômicas, as bactérias sintetizarão as enzimas necessárias para produzir compostos apenas quando não houver outra opção — em outras palavras, quando os compostos estiverem indisponíveis em seu ambiente local.

As bactérias desenvolveram sistemas reguladores que acoplam a expressão dos produtos gênicos a sistemas sensores que detectam o composto relevante no ambiente local de uma bactéria. A regulação das enzimas que participam no metabolismo do açúcar fornece um exemplo. As moléculas de açúcar podem ser fragmentadas para fornecer energia, ou podem ser utilizadas como elementos estruturais para uma grande variedade de compostos orgânicos. Entretanto, existem muitos tipos diferentes de açúcares que as bactérias podem utilizar, incluindo lactose, glicose, galactose e xilose. É necessária uma proteína de importação diferente para que seja possibilitada a entrada de cada um desses açúcares na célula. Além disso, é necessário um conjunto diferente de enzimas para processar cada um dos açúcares. Se uma célula precisasse sintetizar simultaneamente todas as enzimas de que ela pudesse vir a necessitar, a célula gastaria muito mais energia e materiais para produzir as enzimas do que obteria a partir da fragmentação das esperadas fontes de carbono. A célula planejou mecanismos para desligar (reprimir) a transcrição de todos os genes que codificam enzimas não necessárias em uma determinada ocasião e para ligar (ativar) aqueles genes que codificam enzimas necessárias. Por exemplo, se apenas a lactose estiver no ambiente, a célula desligará a transcrição dos genes

que codificam as enzimas necessárias para a importação e o metabolismo da glicose, galactose, xilose e de outros açúcares. Contrariamente, a *E. coli* iniciará a transcrição dos genes que codificam as enzimas necessárias para a importação e o metabolismo da lactose. Resumidamente, as células necessitam de mecanismos que atendam dois critérios:

1. Eles devem ser capazes de reconhecer as condições ambientais nas quais devem ativar ou reprimir a transcrição dos genes relevantes.
2. Eles devem ser capazes de alternar entre ligar e desligar, como um interruptor, a transcrição de cada gene específico ou de um grupo de genes.

Preveremos o atual modelo em relação à regulação da transcrição procariótica e em seguida utilizaremos um exemplo bem-compreendido — a regulação de genes no metabolismo do açúcar lactose — para o seu exame detalhado. Em particular, enfocaremos como esse sistema regulador foi dissecado com a utilização das ferramentas da genética clássica e da biologia molecular.

Base da regulação da transcrição procariótica | Interruptores genéticos

A regulação da transcrição depende principalmente de dois tipos de interações DNA-proteína. Ambas ocorrem próximo do sítio no qual a transcrição gênica tem início.

Uma dessas interações DNA-proteína determina onde a transcrição tem início. O DNA que participa dessa interação é um segmento de DNA denominado **promotor** (Capítulo 8, Seção 8.2), e a proteína que se liga a esse sítio é a RNA polimerase. Quando a RNA polimerase se liga ao DNA promotor, a transcrição pode ter início à distância de algumas poucas bases do sítio promotor. Cada gene deve ter um promotor ou ele não poderá ser transcrito.

O outro tipo de interação DNA-proteína determina se ocorre a transcrição direcionada pelo promotor. Segmentos de DNA próximos ao promotor atuam como sítios de ligação para as proteínas reguladoras sequência-específicas denominadas **ativadores** e **repressores**. Em bactérias, a maior parte dos sítios de ligação para os repressores é denominada **operador**. Em relação a alguns genes,

uma proteína ativadora deve se ligar ao seu sítio-alvo no DNA como um pré-requisito necessário para que a transcrição seja iniciada. Os referidos casos por vezes são denominados *regulação positiva*, tendo em vista que é necessária a *presença* da proteína ligada para a transcrição (Figura 11.2). Em relação a outros genes, deve ser evitada a ligação de uma proteína repressora ao seu sítio-alvo como um pré-requisito necessário para que a transcrição seja iniciada. Os referidos casos por vezes são denominados *regulação negativa*, tendo em vista que a *ausência* do repressor ligado possibilita que a transcrição seja iniciada.

Como os ativadores e os repressores regulam a transcrição? Com frequência, uma proteína ativadora ligada ao DNA auxilia fisicamente a ancorar a RNA polimerase ao seu promotor próximo, de modo que a polimerase possa iniciar a transcrição. Uma proteína repressora ligada ao DNA atua tipicamente ao interferir fisicamente com a ligação da RNA polimerase ao seu promotor (bloqueando o início da transcrição) ou ao impedir a movimentação da RNA polimerase ao longo da cadeia de DNA (bloqueando a transcrição). Em conjunto, essas proteínas reguladoras e seus sítios de ligação constituem **interruptores genéticos** que controlam as alterações eficientes na expressão gênica que ocorrem em resposta às condições ambientais.

> CONCEITO-CHAVE Os interruptores genéticos controlam a transcrição gênica. A função liga/desliga dos interruptores depende das interações de diversas proteínas com seus sítios de ligação no DNA. A RNA polimerase interage com o promotor para iniciar a transcrição. As proteínas ativadoras ou repressoras se ligam a sítios na vizinhança do promotor para controlar sua acessibilidade à RNA polimerase.

Ambas as proteínas ativadoras e repressoras devem ser capazes de reconhecer quando as condições ambientais são apropriadas para as suas ações e agir de acordo. Portanto, para que as proteínas ativadoras ou repressoras realizem o seu trabalho, cada uma deve ser capaz de existir em dois estados: um que possa se ligar a seu DNA-alvo e outro que não possa. O estado de ligação deve ser

apropriado para o conjunto de condições fisiológicas presentes na célula e em seu ambiente. Em relação a muitas proteínas reguladoras, a ligação do DNA é efetuada por meio da interação de dois sítios diferentes na estrutura tridimensional da proteína. Um sítio é o **domínio de ligação ao DNA**. O outro sítio, o **sítio alostérico**, atua como um sensor que configura o domínio de ligação ao DNA em um de dois modos: funcional ou não funcional. O sítio alostérico interage com pequenas moléculas denominadas *efetores alostéricos*.

No metabolismo da lactose, na realidade é um isômero do açúcar lactose (denominado alolactose) que é um efetor alostérico: o açúcar se liga a uma proteína reguladora que inibe a expressão dos genes necessários para o metabolismo da lactose. Em geral, um **efetor alostérico** se liga ao sítio alostérico da proteína reguladora de tal modo a alterar a sua atividade. Nesse caso, a alolactose altera a forma e a estrutura do domínio de ligação ao DNA de uma proteína reguladora. Algumas proteínas ativadoras ou repressoras devem se ligar aos seus efetores alostéricos antes que possam se ligar ao DNA. Outras conseguem se ligar ao DNA somente na ausência de seus efetores alostéricos. Duas dessas situações estão demonstradas na Figura 11.3.

> CONCEITO-CHAVE Os efetores alostéricos controlam a capacidade das proteínas ativadoras ou repressoras de se ligarem aos seus sítios do DNA-alvo.

Primeira observação do circuito regulador lac

O trabalho pioneiro de François Jacob e Jacques Monod na década de 1950 demonstrou como o metabolismo da lactose é regulado geneticamente. Examinaremos o sistema sob duas condições: a presença e a ausência de lactose. A Figura 11.4 é uma visão simplificada dos componentes desse sistema. O elenco de personagens para a regulação do óperon *lac* inclui genes codificadores de proteínas e sítios no DNA que são alvos para proteínas de ligação ao DNA.

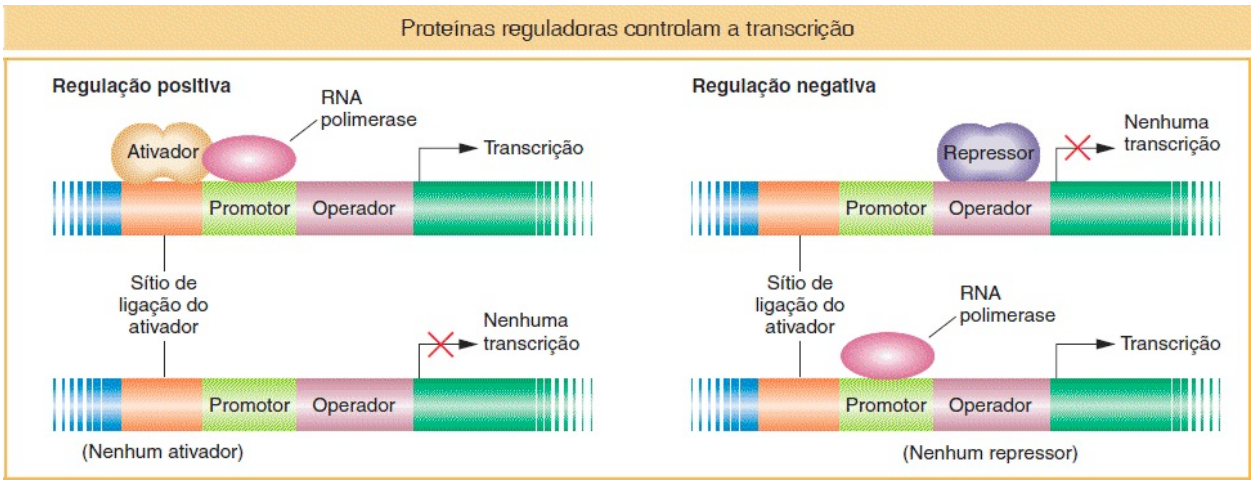


FIGURA 11.2 A ligação de proteínas reguladoras pode ativar ou bloquear a transcrição.

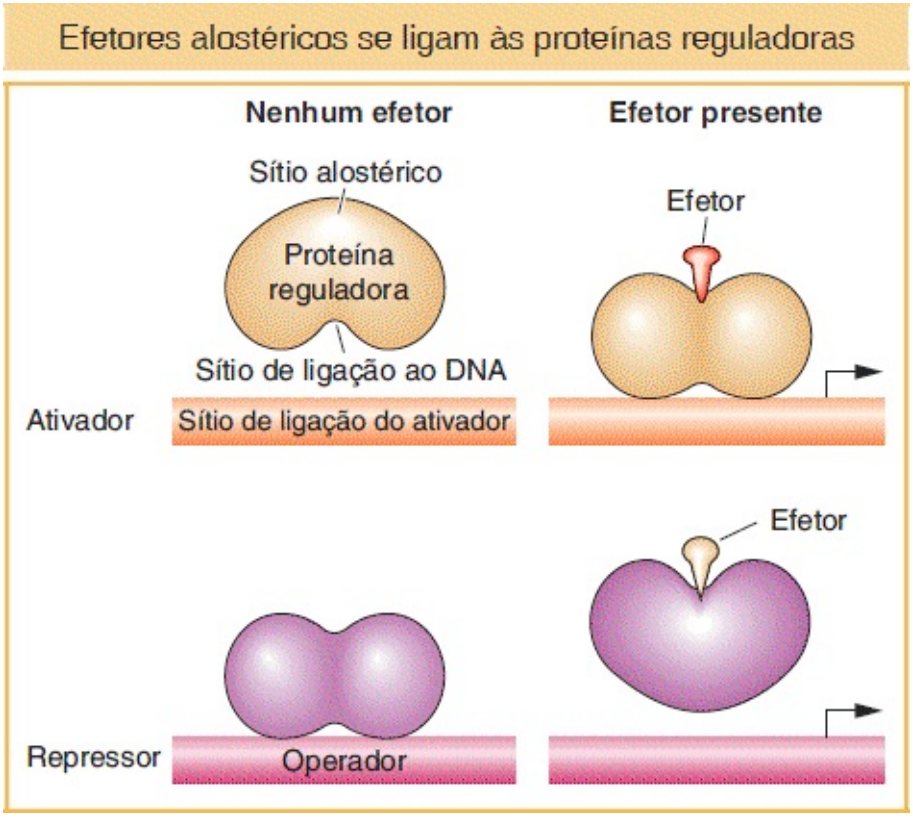


FIGURA 11.3 Efetores alostéricos controlam a capacidade de ligação das proteínas ativadoras ou repressoras a seus sítios-alvo no DNA.

Genes estruturais *lac*. O metabolismo da lactose necessita de duas enzimas: (1) uma permease para transportar a lactose para dentro da célula e (2) β-

galactosidase para modificar a lactose em alolactose e clivar a molécula de lactose para produzir glicose e galactose (Figura 11.5). As estruturas das proteínas β-galactosidase e permease são codificadas por duas sequências adjacentes, *Z* e *Y*, respectivamente. Uma terceira sequência contígua codifica uma enzima adicional, denominada *transacetilase*, que não é necessária para o metabolismo da lactose. Denominaremos *Z*, *Y* e *A genes estruturais* — em outras palavras, segmentos que codificam proteínas — enquanto reservaremos o julgamento sobre essa categorização até posteriormente. Enfocaremos principalmente os genes *Z* e *Y*. Todos os três genes são transcritos em uma única molécula de RNA mensageiro. A regulação da produção desse mRNA coordena a síntese de todas as três enzimas. Ou seja, todas as três enzimas são sintetizadas ou nenhuma o é. Diz-se que os genes cuja transcrição é controlada por um meio comum são **genes controlados coordenadamente**.

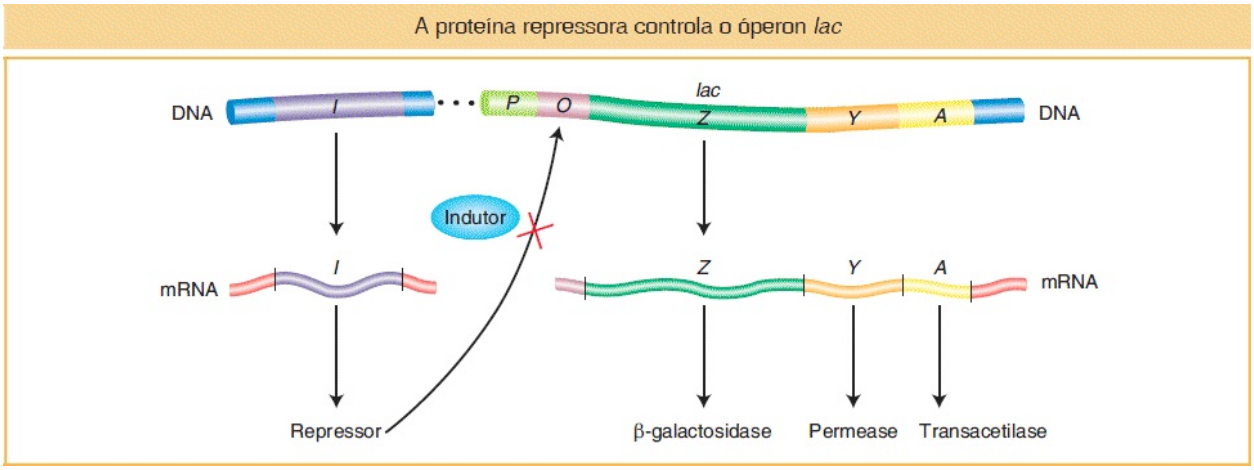


FIGURA 11.4 Um modelo simplificado do óperon *lac*. A expressão coordenada dos genes *Z*, *Y* e *A* está sob o controle negativo do produto do gene *I*, o repressor. Quando o indutor se liga ao repressor, o óperon é totalmente expresso.

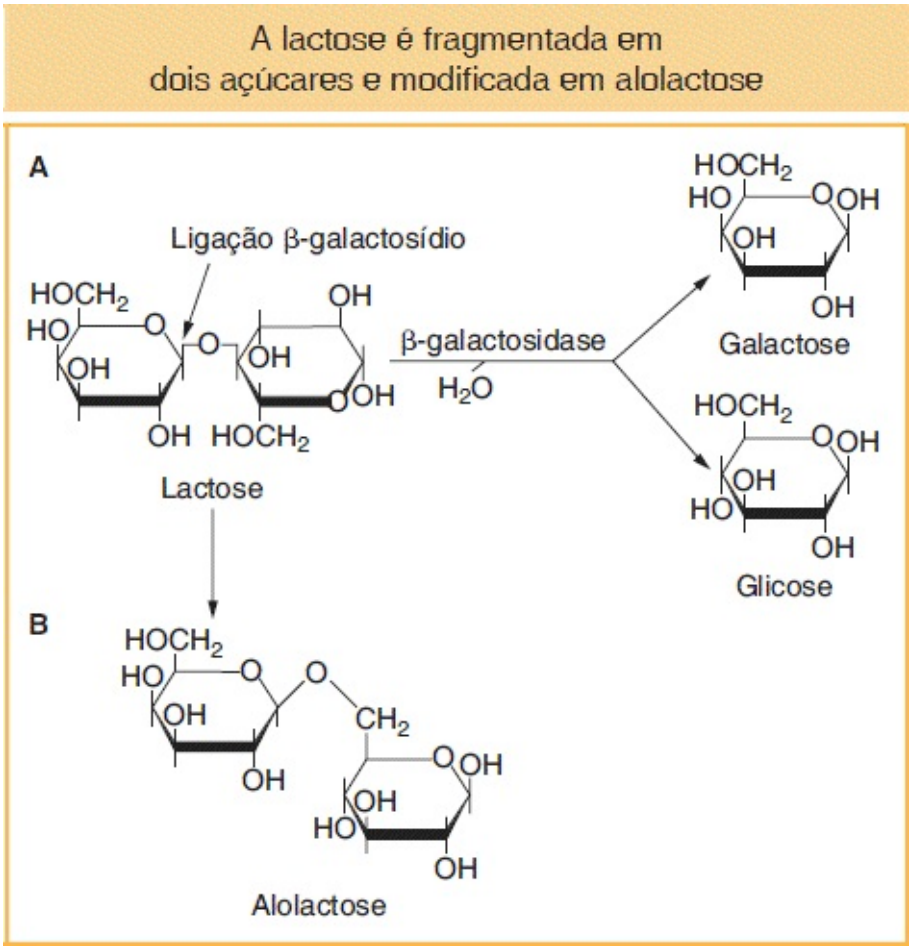


FIGURA 11.5 O metabolismo da lactose. **A.** A enzima β-galactosidase catalisa uma reação na qual é adicionada água à ligação de β-galactosidase para fragmentar a lactose em moléculas separadas de glicose e galactose. **B.** A enzima também modifica uma proporção menor da lactose em alolactose, que atua como um indutor do óperon *lac*.

> CONCEITO-CHAVE Se os genes codificadores de proteínas constituem uma unidade de transcrição única, a expressão de todos esses genes será regulada coordenadamente.

Componentes reguladores do sistema *lac*. Os componentes reguladores-chave do sistema metabólico da lactose incluem um gene que codifica uma proteína reguladora de transcrição e dois sítios de ligação no DNA: um sítio para a proteína reguladora e outro sítio para a RNA polimerase.

1. *O gene para o repressor Lac.* Um quarto gene (além dos genes estruturais *Z, Y* e *A*), o gene *I*, codifica a proteína repressora Lac. Ele é assim denominado em virtude de conseguir bloquear a expressão dos genes *Z, Y* e *A*. Ocorre que o gene *I* está mapeado próximo dos genes *Z, Y* e *A*, mas essa proximidade não é importante para a sua função, tendo em vista que ele codifica uma proteína difundível.
2. *O sítio promotor lac.* O promotor (*P*) é o sítio no DNA ao qual a RNA polimerase se liga para iniciar a transcrição dos genes estruturais *lac* (*Z, Y* e *A*).
3. *O sítio operador lac.* O operador (*O*) é o sítio no DNA ao qual o repressor Lac se liga. Ele está localizado entre o promotor e o gene *Z*, próximo do ponto no qual a transcrição do mRNA multigênico tem início.

Indução do sistema *lac*. Os segmentos *P, O, Z, Y* e *A* (demonstrados na Figura 11.6) em conjunto constituem um **óperon**, definido como um segmento de DNA que codifica um mRNA multigênico, bem como um promotor comum adjacente e uma região reguladora. O gene *lacI*, que codifica o repressor Lac, *não* é considerado parte do próprio óperon *lac*, mas a interação do repressor Lac com o sítio operador *lac* é crucial para a adequada regulação do óperon *lac*. O repressor Lac apresenta um sítio de *ligação ao DNA* que consegue reconhecer a sequência do DNA operador e um *sítio alostérico* que liga a alolactose ou análogos da lactose que são úteis experimentalmente. O repressor se ligará firmemente apenas ao sítio *O* no DNA próximo dos genes que está controlando e não a outras sequências distribuídas por todo o cromossomo. Por meio da ligação ao operador, o repressor evita a transcrição pela RNA polimerase que se ligou ao sítio promotor adjacente; o óperon *lac* é “desligado”.

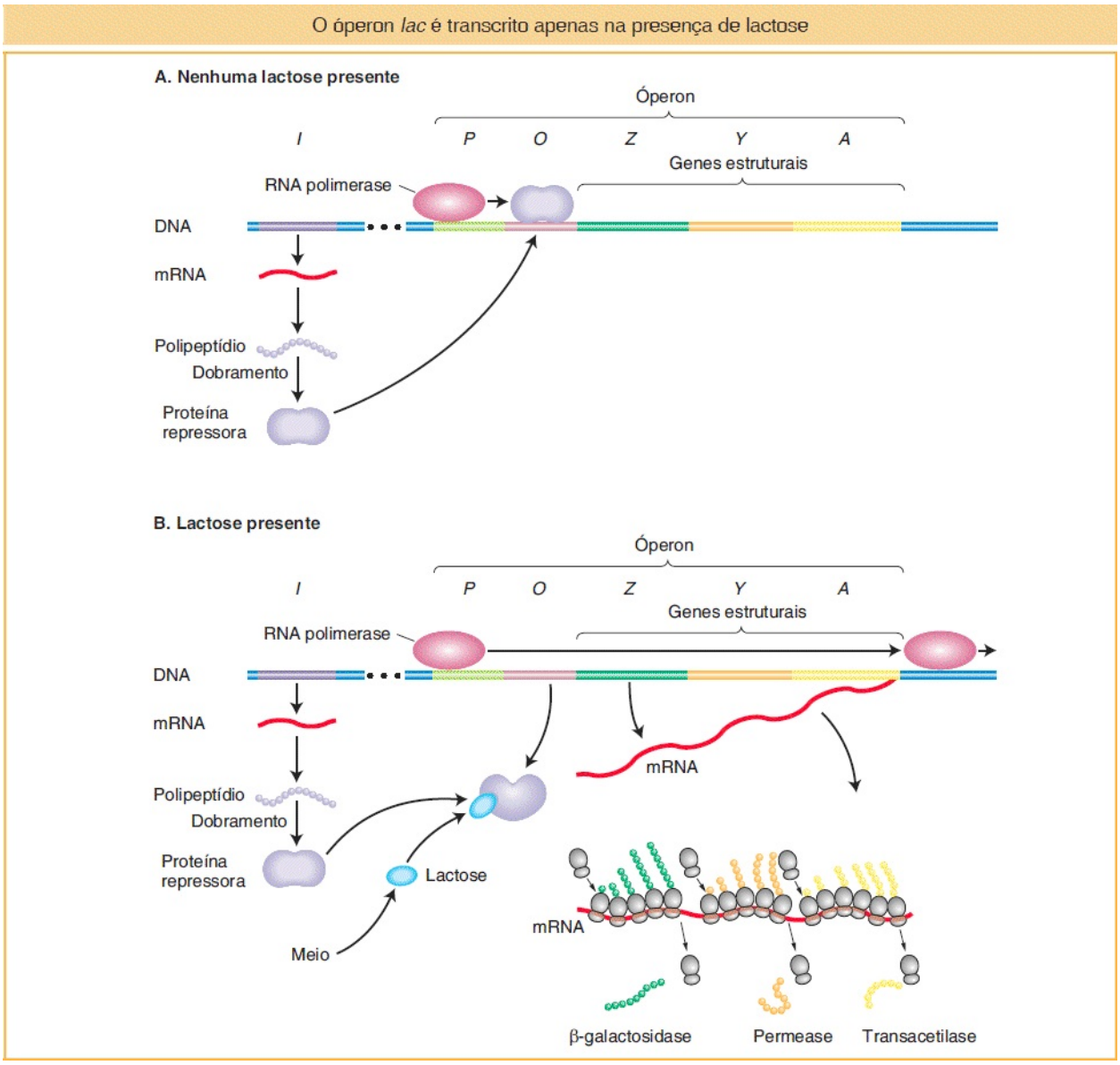


FIGURA 11.6 Regulação do óperon *lac*. O gene *I* produz continuamente o repressor. **A.** Na ausência de lactose, o repressor se liga à região *O* (operador) e bloqueia a transcrição. **B.** A ligação da lactose altera a forma do repressor de modo que o repressor deixa de se ligar ao *O* e sai do DNA. A RNA polimerase é então capaz de transcrever os genes estruturais *Z*, *Y* e *A* e, assim, as três enzimas são produzidas.

Quando a alolactose ou seus análogos se ligam à proteína repressora, a proteína é submetida a uma **transição alostérica**, uma alteração na forma. Essa alteração discreta na forma, por sua vez, altera o sítio de ligação ao DNA de modo que o repressor deixa de apresentar alta afinidade pelo operador. Portanto, em resposta à ligação da alolactose, o repressor sai do DNA, possibilitando que

a RNA polimerase prossiga (transcreva o gene): o óperon *lac* está "ligado". A resposta do repressor à alolactose atende uma exigência em relação ao referido sistema de controle — que a presença da lactose estimula a síntese dos genes necessários para o seu processamento. O alívio da repressão em sistemas tal como o *lac* é denominado **indução**. A alolactose e seus análogos que inativam alostericamente o repressor, levando à expressão dos genes *lac*, são denominados **indutores**.

Resumiremos como o interruptor *lac* atua. Na ausência de um indutor (alolactose ou um análogo), o repressor Lac se liga ao sítio operador *lac* e evita a transcrição do óperon *lac* por meio do bloqueio da movimentação da RNA polimerase. Nesse sentido, o repressor Lac atua como um bloqueio no DNA. Consequentemente, todos os genes estruturais do óperon *lac* (os genes *Z, Y* e *A*) são reprimidos e existem muito poucas moléculas de β-galactosidase, permease ou transacetilase na célula. Contrariamente, quando um indutor está presente, ele se liga ao sítio alostérico de cada subunidade do repressor Lac, inativando, assim, o sítio que se liga ao operador. O repressor Lac sai do DNA, possibilitando que tenha início a transcrição dos genes estruturais do óperon *lac*. As enzimas β-galactosidase, permease e transcetilase agora aparecem na célula de modo coordenado. Assim, quando a lactose está presente no ambiente de uma célula bacteriana, a célula produz as enzimas necessárias para a sua metabolização. Mas quando nenhuma lactose está presente, os recursos não são gastos.

11.2 Descoberta do sistema lac | Controle negativo

Para estudar a regulação gênica, idealmente necessitamos de três ingredientes: um ensaio bioquímico que nos possibilite medir a quantidade de mRNA ou de proteína expressa, ou de ambos, condições confiáveis nas quais os níveis de expressão sejam diferentes em um genótipo do tipo selvagem e mutações genéticas que perturbem os níveis de expressão. Em outras palavras, precisamos de um modo para descrever a regulação do gene do tipo selvagem e necessitamos

de mutações que possam romper o processo regulador do tipo selvagem. Com esses elementos em mãos, podemos analisar a expressão em genótipos mutantes, tratando as mutações de modo único e combinado, para desvendar qualquer tipo de evento de regulação gênica. A aplicação clássica dessa abordagem foi utilizada por Jacob e Monod, que realizaram os estudos definitivos da regulação gênica bacteriana.

Jacob e Monod utilizaram o sistema do metabolismo da lactose de *E. coli* (ver Figura 11.4) para dissecar geneticamente o processo de indução enzimática — ou seja, o aparecimento de uma enzima específica apenas na presença de seus substratos. Esse fenômeno havia sido observado em bactérias por muitos anos, mas como uma célula possivelmente poderia "saber" precisamente quais enzimas devem ser sintetizadas? Como um substrato em particular poderia induzir o aparecimento de uma enzima específica?

No sistema *lac*, a presença de lactose faz com que as células produzam mais de 1.000 vezes a quantidade da enzima β-galactosidase que produzem quando cultivadas sem lactose. Que papel a lactose desempenha no fenômeno da indução? Uma ideia era que a lactose estava simplesmente ativando um precursor de β-galactosidase que havia se acumulado na célula. Entretanto, quando Monod e colaboradores acompanharam o destino de aminoácidos marcados radioativamente adicionados ao cultivo das células antes ou após a adição de um indutor, eles observaram que a indução resultava na síntese de novas moléculas de enzimas, conforme indicado pela presença de aminoácidos radioativos nas enzimas. Essas novas moléculas poderiam ser detectadas tão precocemente quanto três minutos após a adição de um indutor. Além disso, a retirada da lactose ocasionou uma interrupção abrupta na síntese da nova enzima. Portanto, tornou-se claro que a célula apresenta um mecanismo rápido e eficaz para ligar e desligar a expressão gênica em resposta a sinais ambientais.

Genes controlados juntos

Quando Jacob e Monod induziram a β-galactosidase, eles observaram que também induziram a enzima permease, que é necessária para transportar a lactose para dentro da célula. A análise de mutantes indicou que cada enzima era

codificada por um gene diferente. A enzima transacetilase (com uma função dispensável e ainda desconhecida) também foi induzida juntamente com a β-galactosidase e a permease e posteriormente foi demonstrado que ela é codificada por um gene em separado. Portanto, Jacob e Monod puderam identificar três genes controlados de modo coordenado. O mapeamento de recombinação demonstrou que os genes *Z*, *Y* e *A* estavam ligados de modo muito próximo no cromossomo.

Evidências genéticas do operador e do repressor

Agora chegamos ao centro do trabalho de Jacob e Monod: como eles deduziram os mecanismos de regulação gênica no sistema *lac*? Sua estratégia foi uma abordagem genética clássica: examinar as consequências fisiológicas das mutações. Portanto, eles induziram mutações nos genes estruturais e nos elementos reguladores do óperon *lac*. Conforme veremos, as propriedades das mutações nesses diferentes componentes do óperon *lac* são consideravelmente diferentes, e forneceram indicações importantes para Jacob e Monod.

Os indutores naturais, tais como a alolactose, não são ideais para esses experimentos, tendo em vista que são degradados pela β-galactosidase. A concentração do indutor diminui durante o experimento e, assim, as medições da indução enzimática se tornam consideravelmente complicadas. Em vez disso, para os referidos experimentos, Jacob e Monod utilizaram indutores sintéticos, tais como o isopropil-β-D-tiogalactosídeo (IPTG; Figura 11.7). O IPTG não é hidrolisado pela β-galactosidase, mas ainda induz a expressão da enzima β-galactosidase.

Jacob e Monod observaram que diversas classes diferentes de mutações podem alterar a expressão dos genes estruturais do óperon *lac*. Eles estavam interessados na avaliação das interações dos novos alelos, tais como quais alelos exibiam dominância. Mas para realizar os referidos testes, são necessários diploides, e bactérias são haploides. Entretanto, Jacob e Monod foram capazes de produzir bactérias que são parcialmente diploides por meio da inserção de fatores F′ (ver Capítulo 5) que carregam a região *lac* do genoma. Eles então conseguiram criar linhagens que eram heterozigotas em relação a mutações *lac*

selecionadas. Esses **diploides parciais** possibilitaram que Jacob e Monod distinguissem mutações no sítio do DNA regulador (o operador *lac*) a partir de mutações na proteína reguladora (o repressor Lac codificado pelo gene *I*).

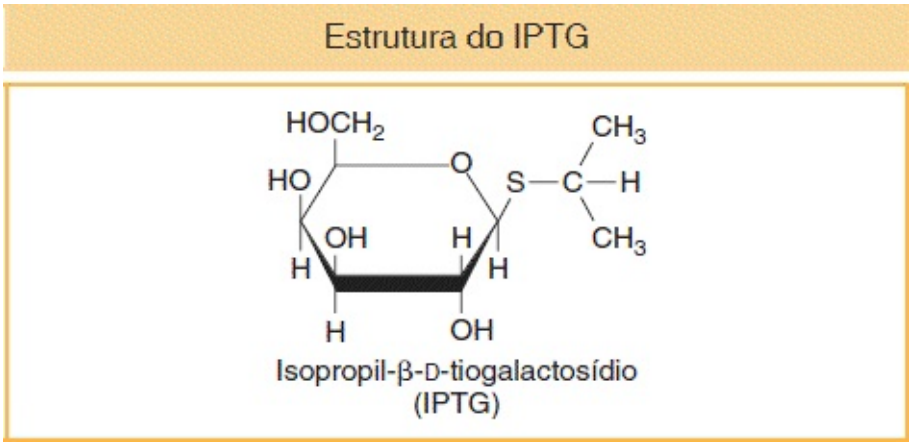


FIGURA 11.7 O IPTG é um indutor do óperon *lac*.

Iniciamos por meio do exame de mutações que inativam os genes estruturais para β-galactosidase e permease (respectivamente designados Z^- e Y^-). A primeira coisa que aprendemos é que Z^- e Y^- são recessivos em relação aos seus respectivos alelos do tipo selvagem (Z^+ e Y^+). Por exemplo, a linhagem 2 na Tabela 11.1 pode ser induzida a sintetizar β-galactosidase (assim como a linhagem 1 haploide do tipo selvagem na tabela) ainda que seja heterozigota em relação aos alelos *Z* mutante e selvagem. Isso demonstra que o alelo Z^+ é dominante em relação ao seu correspondente Z^-.

Jacob e Monod primeiramente identificaram duas classes de mutações reguladoras, denominadas O^C e I^-. Essas foram denominadas **mutações constitutivas**, tendo em vista que causaram a expressão dos genes estruturais do óperon *lac* independentemente da presença do indutor. Jacob e Monod identificaram a existência do operador com base em sua análise das mutações O^C. Essas mutações tornam o operador incapaz de se ligar ao repressor; elas danificam o interruptor de tal modo que o operador está sempre "ligado" (Tabela 11.1, linhagem 3). De importância, os efeitos constitutivos das mutações O^C foram restritos somente àqueles genes estruturais *lac no mesmo cromossomo* da mutação

O^C. Por isso, o operador mutante foi denominado de **ação cis**, conforme demonstrado pelo fenótipo da linhagem 4 na Tabela 11.1. Aqui, tendo em vista que o gene da permease do tipo selvagem (Y^+) é cis em relação ao operador do tipo selvagem, a permease é expressa apenas quando a lactose ou um análogo estão presentes. Contrariamente, o gene da β-galactosidase do tipo selvagem (Z^+) é cis em relação ao operador mutante O^C; portanto, a β-galactosidase é expressa constitutivamente. Essa propriedade incomum da ação cis sugeriu que o operador é um segmento de DNA que influencia apenas a expressão dos genes estruturais ligados a ele (Figura 11.8). Portanto, o operador atua simplesmente como um sítio de ligação à proteína e *não faz* produto gênico.

Jacob e Monod realizaram testes genéticos comparáveis com as mutações I^- (Tabela 11.2). Uma comparação da linhagem I^+ do tipo selvagem indutível (linhagem 1) com linhagens I^- demonstra que as mutações I^- são constitutivas (linhagem 2). Ou seja, elas causam a expressão dos genes estruturais o tempo todo. A linhagem 3 demonstra que o fenótipo indutível de I^+ é dominante em relação ao fenótipo constitutivo de I^-. Esse achado demonstrou a Jacob e Monod que a quantidade de proteína do tipo selvagem codificada por uma cópia do gene é suficiente para regular ambas as cópias do operador em uma célula diploide. Mais significativamente, a linhagem 4 demonstrou a eles que o produto do gene I^+ é **transatuante**, o que significa que o produto do gene consegue regular *todos* os genes estruturais do óperon *lac*, estejam eles localizados na mesma molécula de DNA ou em moléculas diferentes (em cis ou em trans, respectivamente). Contrariamente ao operador, o gene *I* se comporta como um gene-padrão codificador de proteína. O produto proteico do gene *I* é capaz de se difundir por uma célula e atuar sobre ambos os operadores no diploide parcial (Figura 11.9).

Tabela 11.1 Síntese de β-galactosidase e permease em mutantes operadores haploides e diploides heterozigotos.

	β-galactosidase (Z)	***Permease (Y)***
Não	**Não**	

Linhagem	Genótipo	β-galactosidase induzida	Permease Induzida	Transacetilase induzida	Transacetilase Induzida
1	$O^+ Z^+ Y^+$	—	+	—	+
2	$O^+ Z^+ Y^+/F' O^+ Z^- Y^+$	—	+	—	+
3	$O^C Z^+ Y^+$	+	+	+	+
4	$O^+ Z^- Y^+/F' O^C Z^+ Y^-$	+	+	—	+

Nota: as bactérias foram cultivadas em glicerol (nenhuma glicose presente) com e sem o indutor IPTG. Presença e ausência da enzima são indicadas por + ou —, respectivamente. Todas as linhagens são I^+.

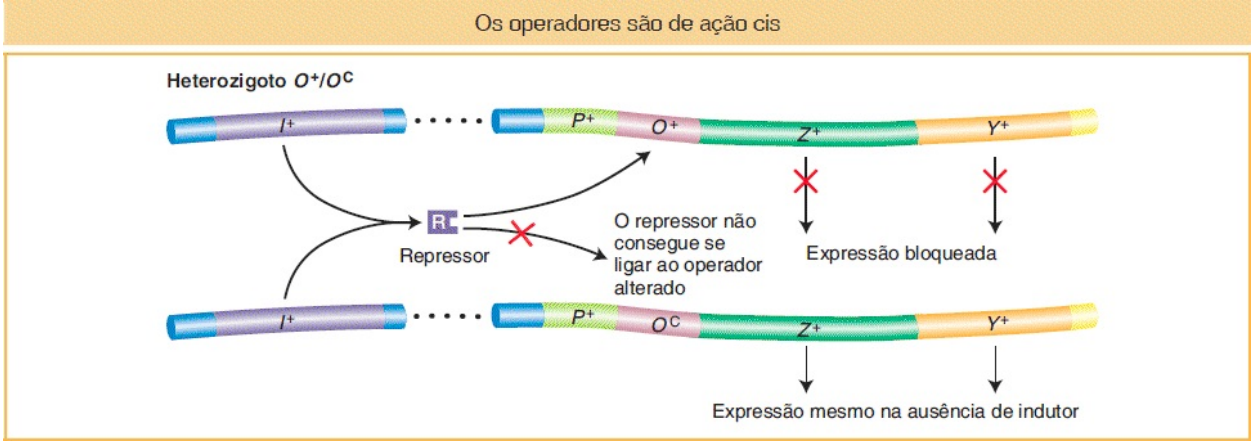


FIGURA 11.8 Os heterozigotos O^+/O^C demonstram que os operadores são de ação cis. Tendo em vista que um repressor não consegue se ligar aos operadores O^C, os genes estruturais *lac* ligados a um operador O^C são expressos mesmo na ausência de um indutor. Entretanto, os genes *lac* adjacentes a um operador O^+ ainda estão sujeitos à repressão.

> CONCEITO-CHAVE As mutações no operador revelam que um referido sítio é de ação cis; ou seja, ele regula a expressão de uma unidade de transcrição adjacente na mesma molécula de DNA. Contrariamente, mutações no gene codificador de uma proteína repressora revelam que essa proteína é transatuante; ou seja, ela pode atuar sobre qualquer cópia do DNA-alvo.

Evidências genéticas em relação à alosteria

Finalmente, Jacob e Monod foram capazes de demonstrar alosteria por meio da análise de outra classe de mutações repressoras. Relembre que o repressor Lac inibe a transcrição do óperon *lac* na ausência de um indutor, mas possibilita a transcrição quando o indutor está presente. Essa regulação é realizada por meio de um segundo sítio na proteína repressora, o sítio alostérico, que se liga ao indutor. Quando ligado ao indutor, o repressor é submetido a uma alteração na estrutura geral de tal modo que seu sítio de ligação ao DNA deixa de funcionar.

Jacob e Monod isolaram outra classe de mutações repressoras, denominadas mutações super-repressoras (I^S). As mutações I^S causam a persistência da repressão até mesmo na presença de um indutor (compare a linhagem 2 na Tabela 11.3 com a linhagem 1 do tipo selvagem indutível). Contrariamente às mutações I^-, as mutações I^S são dominantes em relação a I^+ (ver Tabela 11.3, linhagem 3). Essa observação-chave levou Jacob e Monod a especularem que as mutações I^S alteram o sítio alostérico de tal modo que ele não consegue mais se ligar a um indutor. Por consequência, a proteína repressora codificada por I^S se liga continuamente ao operador — evitando a transcrição do óperon *lac* mesmo quando o indutor está presente na célula. Com base nisso, podemos ver o motivo de I^S ser dominante em relação a I^+. A proteína I^S mutante se ligará a ambos os operadores na célula, mesmo na presença de um indutor e independentemente do

fato de que a proteína codificada por I^+ pode estar presente na mesma célula (Figura 11.10).

Tabela 11.2 Síntese de β-galactosidase e permease em linhagens haploides e diploides heterozigotas que carregam I^+ e I^-.

		β-galactosidase (Z)		***Permease (Y)***	
Linhagem	**Genótipo**	**Não induzida**	**Induzida**	**Não induzida**	**Induzida**
1	$I^+ Z^+ Y^+$	—	+	—	+
2	$I^- Z^+ Y^+$	+	+	+	+
3	$I^+ Z^-$ Y^+/F′ I^- $Z^+ Y^+$	—	+	—	+
4	$I^- Z^-$ Y^+/F′ I^+ $Z^+ Y^-$	—	+	—	+

Nota: as bactérias foram cultivadas em glicerol (nenhuma glicose presente) e induzidas com IPTG. A presença do nível máximo da enzima é indicada por um sinal de mais; a ausência ou o nível muito baixo de uma enzima é indicado por um

sinal de menos. (Todas as linhagens são O^+).

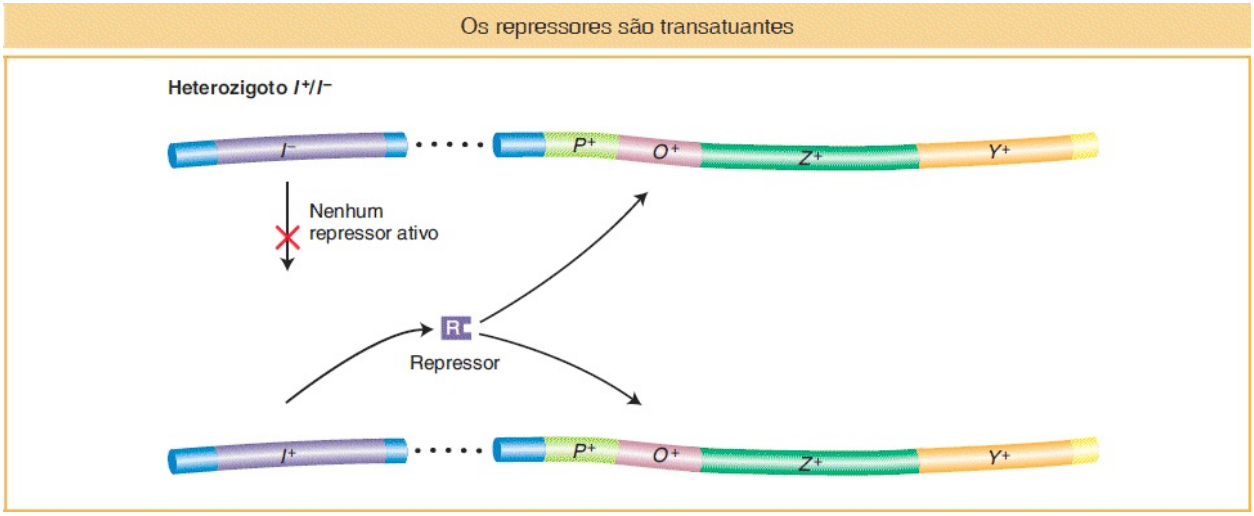


FIGURA 11.9 A natureza recessiva das mutações I^- demonstra que o repressor é transatuante. Embora nenhum repressor ativo seja sintetizado a partir do gene I^-, o gene do tipo selvagem (I^+) proporciona um repressor funcional que se liga a ambos os operadores em uma célula diploide e bloqueia a expressão do óperon *lac* (na ausência de um indutor).

Tabela 11.3 Síntese de β-galactosidase e permease pelo tipo selvagem e por linhagens que carregam alelos diferentes do gene I.

		β-galactosidase (Z)		***Permease (Y)***	
Linhagem	**Genótipo**	**Não induzida**	**Induzida**	**Não induzida**	**Induzida**
1	$I^+ Z^+ Y^+$	—	+	—	+
2	$I^S Z^+ Y^+$	—	—	—	—

3	$I^S\ Z^+Y^+/F'$ $I^+\ Z^+\ Y^+$	—	—	—	—

Nota: as bactérias foram cultivadas em glicerol (sem glicose) com e sem o indutor IPTG. A presença da enzima indicada é representada por +; a ausência ou níveis baixos, por —.

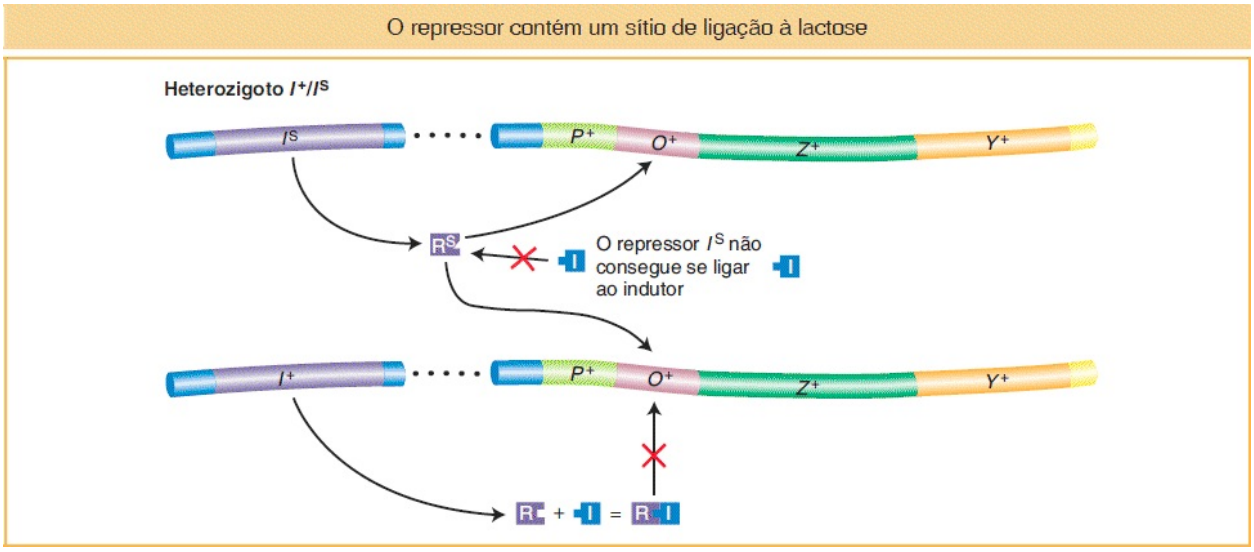


FIGURA 11.10 A dominância da mutação I^S ocorre em virtude da inativação do sítio alostérico no repressor Lac. Em uma célula diploide I^S/I^-, nenhum dos genes estruturais do *lac* é transcrito. O repressor I^S não apresenta um sítio de ligação à alolactose (sítio alostérico) funcional e, portanto, não é inativado por um indutor. Portanto, até na presença de um indutor, o repressor I^S se liga irreversivelmente a todos os operadores em uma célula, bloqueando a transcrição do óperon *lac*.

Análise genética do promotor lac

A análise mutacional também demonstrou que um elemento essencial para a transcrição de *lac* está localizado entre o gene para o repressor *I* e o sítio operador *O*. Esse elemento, denominado *promotor* (*P*), atua como o sítio de iniciação para a transcrição pela RNA polimerase, conforme descrito no Capítulo 8. Existem duas regiões de ligação para a RNA polimerase em um promotor procariótico típico, demonstradas na Figura 11.11 como as duas regiões altamente

conservadas em —35 e —10. As mutações do promotor são de ação cis, pois afetam a transcrição de todos os genes estruturais adjacentes no óperon. Assim como os operadores e outros elementos de ação cis, os promotores são sítios na molécula de DNA que são ligados a proteínas e eles próprios não dão origem a produto proteico.

Caracterização molecular do repressor Lac e do operador *lac*

Walter Gilbert e Benno Müller-Hill forneceram uma demonstração decisiva do sistema *lac* em 1966 por meio do monitoramento da ligação do indutor IPTG marcado radioativamente à proteína repressora purificada. Primeiramente eles demonstraram que o repressor é composto por quatro subunidades idênticas e, portanto, contém quatro sítios de ligação de IPTG (e, portanto, de ligação de alolactose). Em segundo lugar, eles demonstraram que, no tubo de ensaio, a proteína repressora se liga a um DNA que contém o operador e que sai do DNA na presença de IPTG. (Uma descrição mais detalhada sobre como o repressor e outras proteínas de ligação ao DNA atuam é fornecida posteriormente, no fim da Seção 11.6.)

Gilbert e seus colaboradores demonstraram que o repressor pode proteger bases específicas no operador contra reagentes químicos. Essa informação possibilitou que eles isolassem o segmento de DNA que constitui o operador e determinassem a sua sequência. Eles usaram o DNA do óperon ao qual o repressor estava ligado e o trataram com a enzima DNase, que fragmenta o DNA. Foram capazes de recuperar trechos curtos de DNA que haviam sido protegidos da atividade enzimática pela molécula repressora. Esses trechos curtos presumivelmente constituíam a sequência do operador. A sequência de bases de cada filamento foi determinada, e foi demonstrado que cada mutação no operador é uma alteração na sequência (Figura 11.12). Esses resultados demonstraram que o *locus* operador é uma sequência específica de 17 a 25 nucleotídios situados logo antes (em 5′) do gene estrutural *Z*. Eles também demonstraram a incrível especificidade do reconhecimento do repressor–operador, que pode ser rompida pela substituição de uma única base. Quando a sequência de bases no mRNA *lac* (transcrito a partir do óperon *lac*) foi determinada, as primeiras 21 bases na

extremidade de iniciação 5′ comprovaram ser complementares à sequência do operador que Gilbert havia determinado, demonstrando que a sequência operadora é transcrita.

Os resultados desses experimentos forneceram confirmações cruciais sobre o mecanismo da ação repressora formulado por Jacob e Monod.

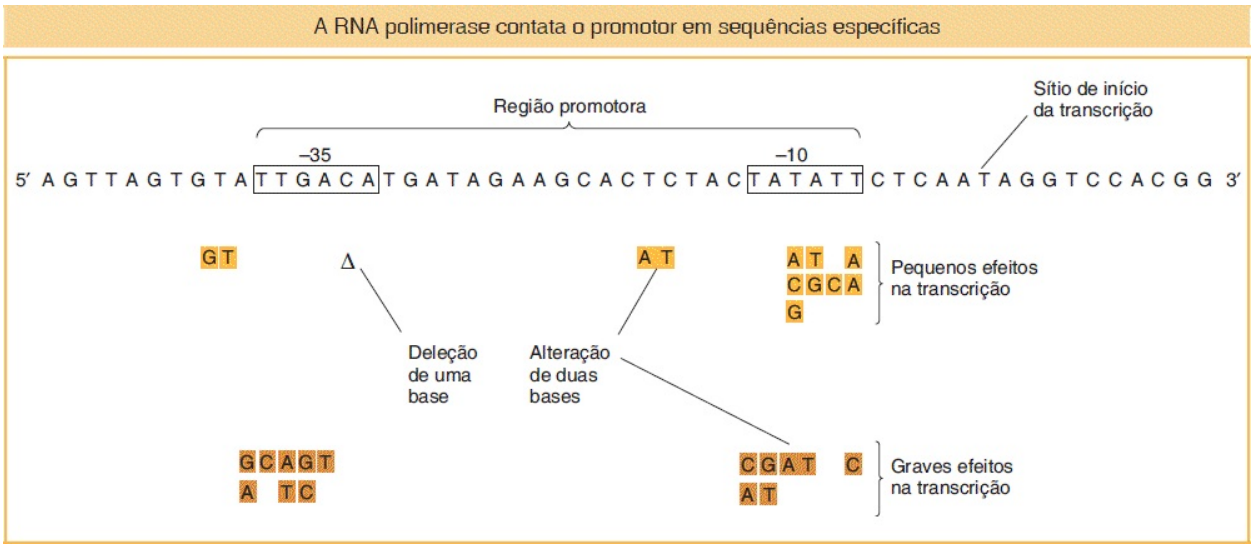


FIGURA 11.11 Sequências de DNA específicas são importantes para a transcrição eficiente dos genes de *E. coli* pela RNA polimerase. Apenas o filamento não molde está demonstrado aqui (ver Figura 8.4). A transcrição prosseguiria da esquerda para a direita (5′ para 3′), e o transcrito de mRNA seria homólogo à sequência demonstrada. As sequências em quadro são altamente conservadas em todos os promotores de *E. coli*, uma indicação do seu papel como sítios de contato no DNA para a ligação da RNA polimerase, e os contatos são realizados com ambos os filamentos (*não demonstrados*). Mutações nessas regiões apresentam efeitos leves (*dourados*) e graves (*marrons*) sobre a transcrição. As mutações podem ser alterações de nucleotídios únicos ou pares de nucleotídios, ou pode ocorrer uma deleção (Δ). (Dados de J. D. Watson, M. Gilman, J. Witkowski e M. Zoller, *Recombinant DNA*, 2nd ed.)

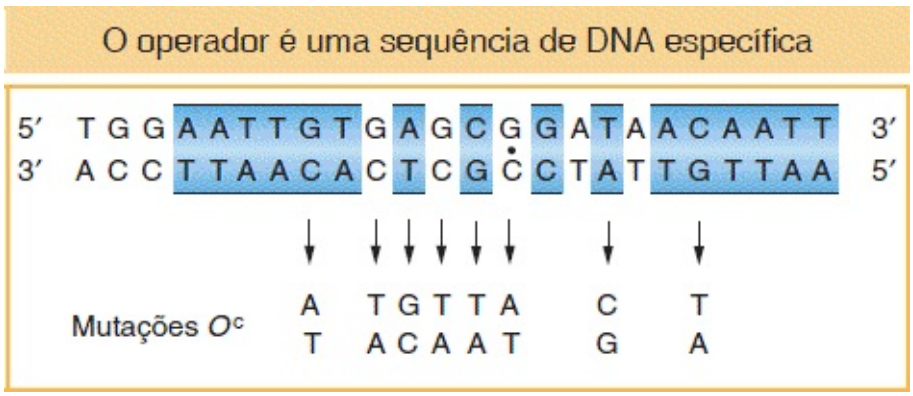


FIGURA 11.12 A sequência de bases do DNA do operador da lactose e as alterações de bases associadas a oito mutações O^C. As regiões de simetria rotacional dupla estão indicadas pela cor e por um ponto em seu eixo de simetria. (Dados de W. Gilbert, A. Maxam e A. Mirzabekov, in N. O. Kjeldgaard e O. Malløe, eds., *Control of Ribosome Synthesis*. Academic Press, 1976.)

11.3 Repressão catabólica do óperon lac | Controle positivo

Durante um longo processo evolutivo, o sistema *lac* existente foi selecionado para operar para a eficiência energética ideal da célula bacteriana. Presumivelmente para maximizar a eficiência energética, duas condições ambientais devem ser atendidas para que as enzimas metabólicas da lactose sejam expressas.

Uma condição é que a lactose deve estar presente no ambiente. Essa condição faz sentido, tendo em vista que seria ineficiente para a célula produzir enzimas metabólicas de lactose se não houver lactose a ser metabolizada. Já vimos que a célula é capaz de responder à presença da lactose por meio da ação de uma proteína repressora.

A outra condição é que a glicose não pode estar presente no ambiente da célula. Tendo em vista que a célula consegue capturar mais energia a partir da fragmentação da glicose do que consegue a partir da fragmentação de outros açúcares, é mais eficiente para a célula metabolizar glicose do que lactose. Portanto, foram desenvolvidos mecanismos que evitam que a célula sintetize as enzimas para o metabolismo da lactose quando tanto lactose quanto glicose estão presentes. A repressão da transcrição dos genes que metabolizam lactose na presença da glicose é um exemplo de **repressão catabólica** (a glicose é um produto da fragmentação, ou um *catabólito*, da lactose). A transcrição dos genes que codificam as proteínas necessárias para o metabolismo de muitos açúcares diferentes é reprimida de modo semelhante na presença de glicose. Veremos que a repressão catabólica atua por meio de uma *proteína ativadora*.

Base da repressão catabólica de lac | Escolha do melhor açúcar a ser metabolizado

Se tanto a lactose quanto glicose estiverem presentes, a síntese de β-galactosidase não é induzida até que toda a glicose tenha sido metabolizada. Portanto, a célula conserva a sua energia metabolizando qualquer glicose existente antes de realizar o processo de gasto energético de criação de um novo maquinário para metabolizar a lactose. Existem múltiplos mecanismos que as bactérias desenvolveram para assegurar a utilização preferencial de uma fonte de carbono e o crescimento ideal. Um mecanismo é excluir a lactose da célula. Um segundo mecanismo é regular a expressão do óperon por meio de catabólitos.

Os resultados de estudos indicam que o produto da fragmentação da glicose evita a ativação do óperon *lac* pela lactose — a repressão catabólica mencionada há pouco. A identidade desse produto da fragmentação ainda é desconhecida. Entretanto, o produto da fragmentação da glicose sabidamente modula o nível de um constituinte celular importante — a **adenosina monofosfato cíclico (cAMP)**. Quando a glicose está presente em altas concentrações, a concentração de cAMP na célula é baixa. Na medida em que a concentração da glicose diminui, a concentração de cAMP da célula aumenta de modo correspondente (Figura 11.13 A). É necessária uma alta concentração de cAMP para a ativação do óperon *lac*. Mutantes que não conseguem converter ATP em cAMP não podem ser induzidos a produzir β-galactosidase, tendo em vista que a concentração de cAMP não é grande o suficiente para ativar o óperon *lac*.

Qual é o papel da cAMP na ativação de *lac*? O estudo de um conjunto diferente de mutantes forneceu uma resposta. Esses mutantes produzem cAMP, mas não conseguem ativar as enzimas Lac, tendo em vista que elas não tem outra proteína, denominada **proteína ativadora de catabólito (CAP)**, codificada pelo gene *crp*. A CAP se liga a uma sequência de DNA específica do óperon *lac* (o sítio de ligação à CAP; ver Figura 11.14 B). A CAP ligada ao DNA em seguida é capaz de interagir fisicamente com a RNA polimerase e aumentar a afinidade da enzima pelo promotor *lac*. Por si própria, a CAP não consegue se ligar ao sítio de ligação à CAP do óperon *lac*. Entretanto, por meio da ligação à cAMP, seu efetor alostérico, a CAP é capaz de se ligar ao sítio de ligação à CAP e ativar a transcrição pela RNA polimerase (Figura 11.13 B). Por meio da inibição da CAP quando a glicose está disponível, o sistema de repressão catabólica assegura que

o óperon *lac* será ativado apenas quando a glicose for escassa.

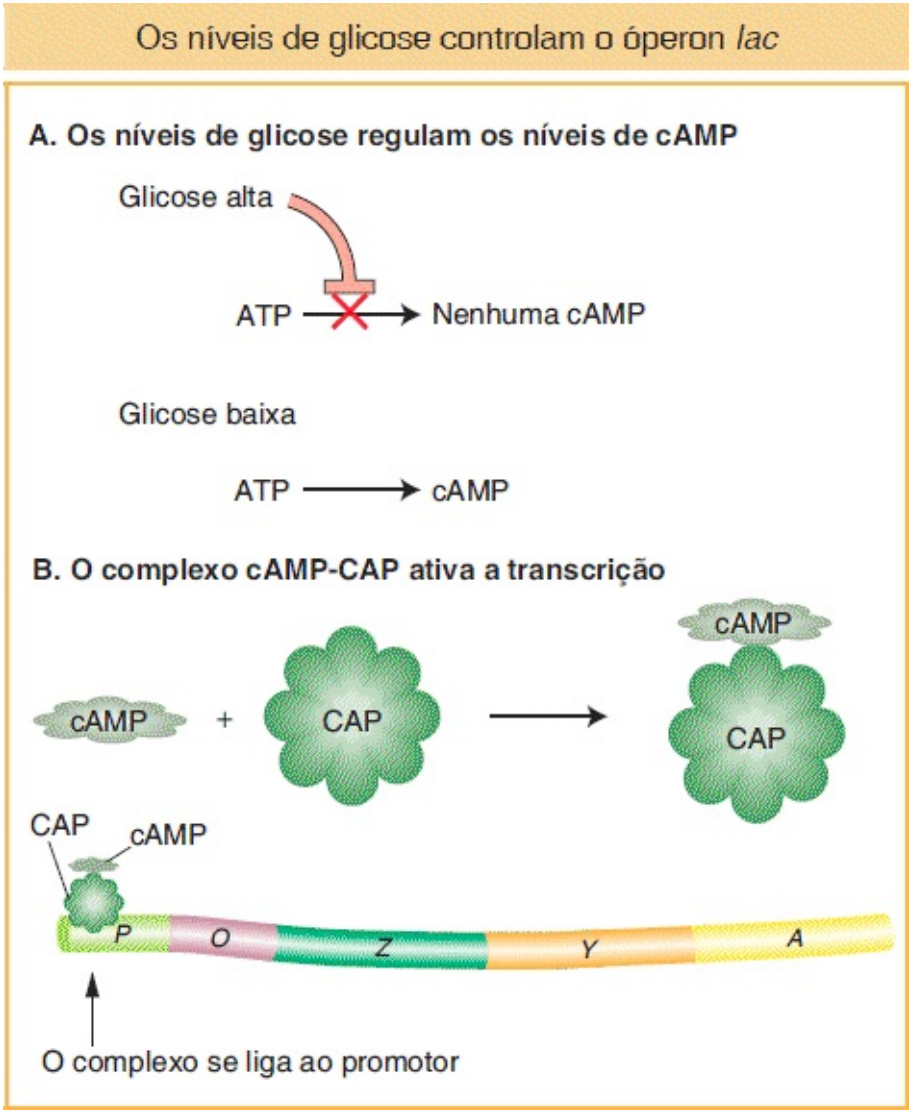


FIGURA 11.13 O controle catabólico do óperon *lac*. **A.** Apenas sob condições de glicose baixa a cAMP (adenosina monofosfato cíclico) é formada. **B.** Quando cAMP está presente, ele forma um complexo com a CAP (proteína ativadora de catabólito) que ativa a transcrição por meio da ligação a uma região no promotor do *lac*.

CONCEITO-CHAVE O óperon *lac* apresenta um nível de controle adicional, de modo que o óperon é inativo na presença de glicose até mesmo se a lactose também estiver presente. Um efetor alostérico, cAMP, liga-se ao ativador CAP para possibilitar a indução do óperon *lac*. Entretanto,

> concentrações altas de catabólitos da glicose inibem a produção de cAMP, falhando, assim, em produzir cAMP-CAP e, portanto, deixando de ativar o óperon *lac*.

Estrutura dos sítios-alvo no DNA

As sequências de DNA às quais o complexo CAP-cAMP se liga (ver Figura 11.14) são diferentes das sequências às quais o repressor Lac se liga. Essas diferenças são a base da especificidade da ligação ao DNA por parte dessas proteínas reguladoras muito diferentes. Uma propriedade que essas sequências apresentam em comum e que é comum a muitos outros sítios de ligação ao DNA é a simetria rotacional dupla. Em outras palavras, se girarmos a sequência de DNA demonstrada na Figura 11.14 em 180° dentro do plano da página, a sequência das bases destacadas dos sítios de ligação será idêntica. Acredita-se que as bases destacadas constituam os importantes sítios de contato para as interações proteína-DNA. Essa simetria rotacional corresponde às simetrias nas proteínas de ligação ao DNA, muitas das quais são compostas por duas ou quatro subunidades idênticas. Consideraremos as estruturas de algumas proteínas de ligação ao DNA posteriormente no capítulo.

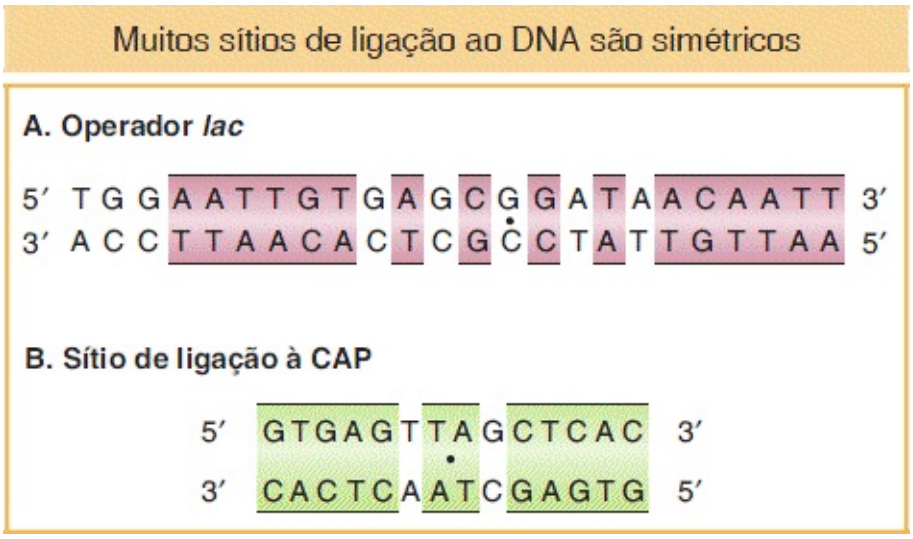


FIGURA 11.14 As sequências de bases do DNA (**A**) do operador *lac*, ao qual o repressor Lac se liga, e (**B**) do sítio de ligação da CAP, ao qual o complexo CAP-cAMP se liga. As sequências que exibem simetria

rotacional dupla estão indicadas pelos *quadros coloridos* e por um *ponto* na parte central da simetria. (*A.* Dados de W. Gilbert, A. Maxam e A. Mirzabekov, in N. O. Kjeldgaard e O. Malløe, eds., *Control of Ribosome Synthesis*. Academic Press, 1976.)

Como a ligação do complexo cAMP-CAP ao óperon promove a ligação da RNA polimerase ao promotor *lac*? Na Figura 11.15, o DNA está dobrado quando a CAP está ligada. Esse dobramento do DNA pode auxiliar a ligação da RNA polimerase ao promotor. Também existem evidências de que a CAP realiza um contato direto com a RNA polimerase, que é importante para o efeito de ativação da CAP. A sequência de bases demonstra que a CAP e a RNA polimerase se ligam de modo diretamente adjacente entre si no promotor *lac* (Figura 11.16).

> CONCEITO-CHAVE Generalizando a partir do modelo do óperon *lac*, podemos imaginar o DNA ocupado por proteínas reguladoras ligadas aos sítios operadores que elas controlam. O padrão exato da ligação dependerá de quais genes estão ligados ou desligados e do fato de ativadores ou repressores regularem óperons em particular.

Resumo do óperon lac

Agora podemos ajustar os sítios de ligação da CAP-cAMP e da RNA polimerase ao modelo detalhado do óperon *lac*, conforme demonstrado na Figura 11.17. A presença de glicose evita o metabolismo da lactose, tendo em vista que um produto da fragmentação da glicose inibe a manutenção dos altos níveis de cAMP necessários para a formação do complexo CAP-cAMP, que por sua vez é necessário para que a RNA polimerase se ligue ao sítio do promotor *lac* (ver Figura 11.17 A e B). Até quando há uma escassez de catabólitos de glicose e formação de CAP-cAMP, o mecanismo para o metabolismo da lactose será implementado apenas se a lactose estiver presente (ver Figura 11.17 C). Apenas duas ou três moléculas de β-galactosidase estão presentes por célula na ausência de lactose ou na presença de lactose e glicose. Essas aumentam para aproximadamente 3.000 moléculas da enzima quando a lactose está presente e a

glicose está ausente. Portanto, a célula conserva a sua energia e seus recursos por meio da produção de enzimas que metabolizam a lactose apenas quando elas são tanto necessárias quanto úteis.

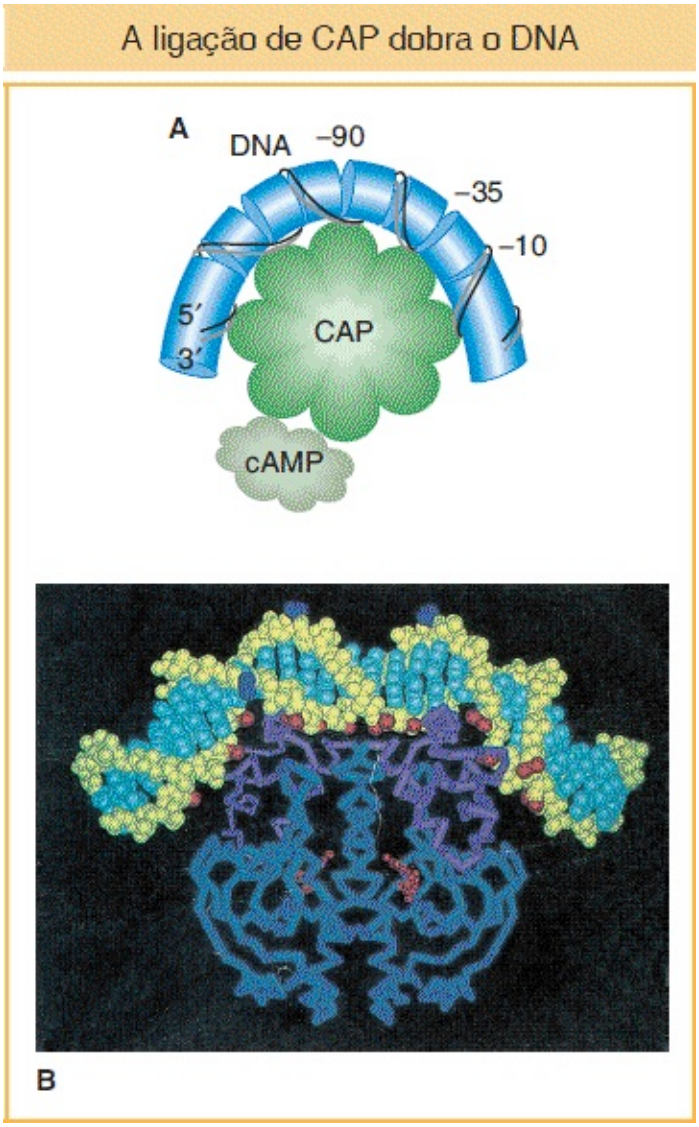


FIGURA 11.15 A. Quando a CAP se liga ao promotor, ela cria uma dobra superior a 90° no DNA. **B.** Imagem derivada da análise estrutural do complexo CAP-DNA. (*B.* De S. Schultz e T. A. Steitz.)

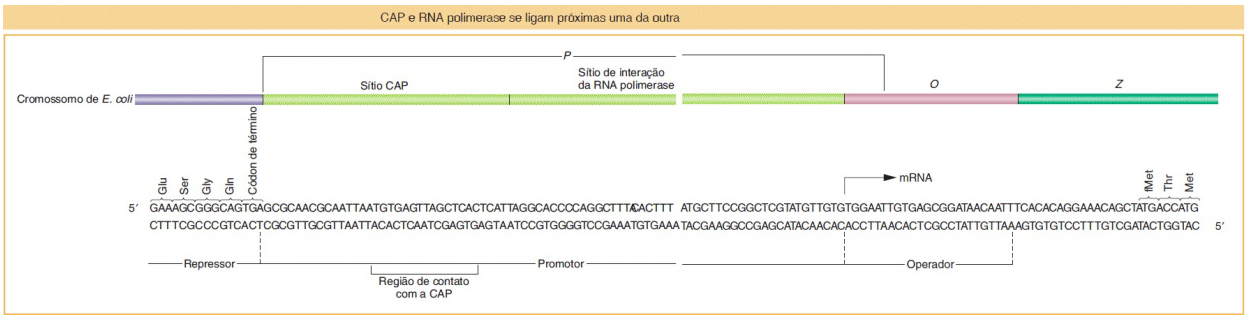


FIGURA 11.16 A região de controle do óperon *lac*. A sequência de bases e os limites genéticos da região de controle do óperon *lac*, com sequências parciais para os genes estruturais. (Dados de R. C. Dickson, J. Abelson, W. M. Barnes e W. S. Reznikoff, “Genetic Regulation: The Lac Control Region”, *Science* 187, 1975, 27.)

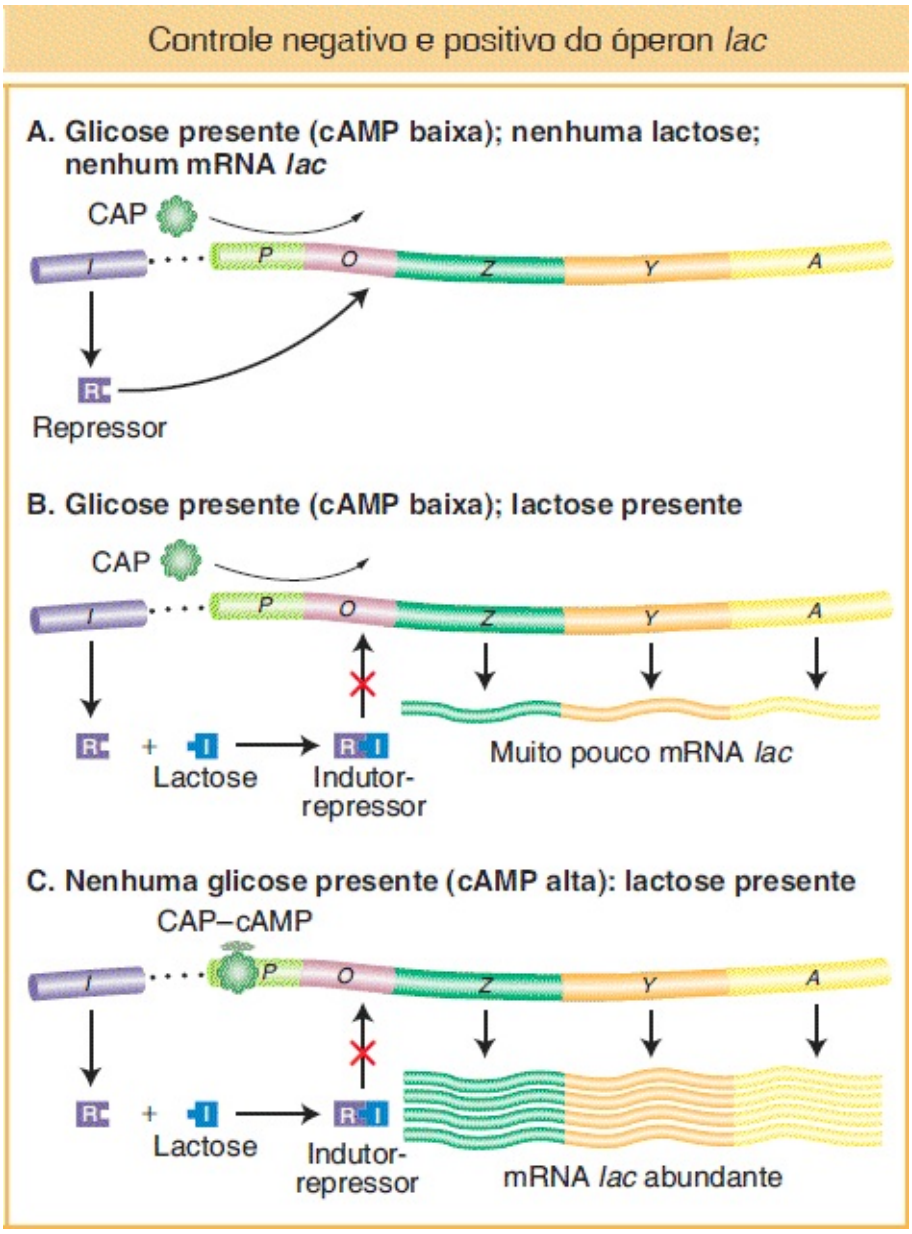


FIGURA 11.17 O óperon *lac* é controlado em conjunto pelo repressor Lac (controle negativo) e pela proteína ativadora de catabólito (CAP; controle positivo). Grandes quantidades de mRNA são produzidas apenas quando a lactose está presente para inativar o repressor, e baixos níveis de glicose promovem a formação do complexo CAP-cAMP, que regula positivamente a transcrição.

O controle do indutor-repressor do óperon *lac* é um exemplo de repressão, ou **controle negativo**, no qual a expressão normalmente está bloqueada. Contrariamente, o sistema de CAP-cAMP é um exemplo de ativação, ou **controle positivo**, tendo em vista que ele atua como um sinal que ativa a expressão —

nesse caso, o sinal de ativação é a interação do complexo CAP-cAMP com o sítio de ligação da CAP ao DNA. A Figura 11.18 resume esses dois tipos básicos de sistemas de controle.

> CONCEITO-CHAVE O óperon *lac* é um agrupamento de genes estruturais que especificam enzimas que participam no metabolismo da lactose. Esses genes são controlados pelas ações coordenadas das regiões do promotor e operador de ação cis. A atividade dessas regiões, por sua vez, é determinada por moléculas repressoras e ativadoras especificadas por genes reguladores em separado.

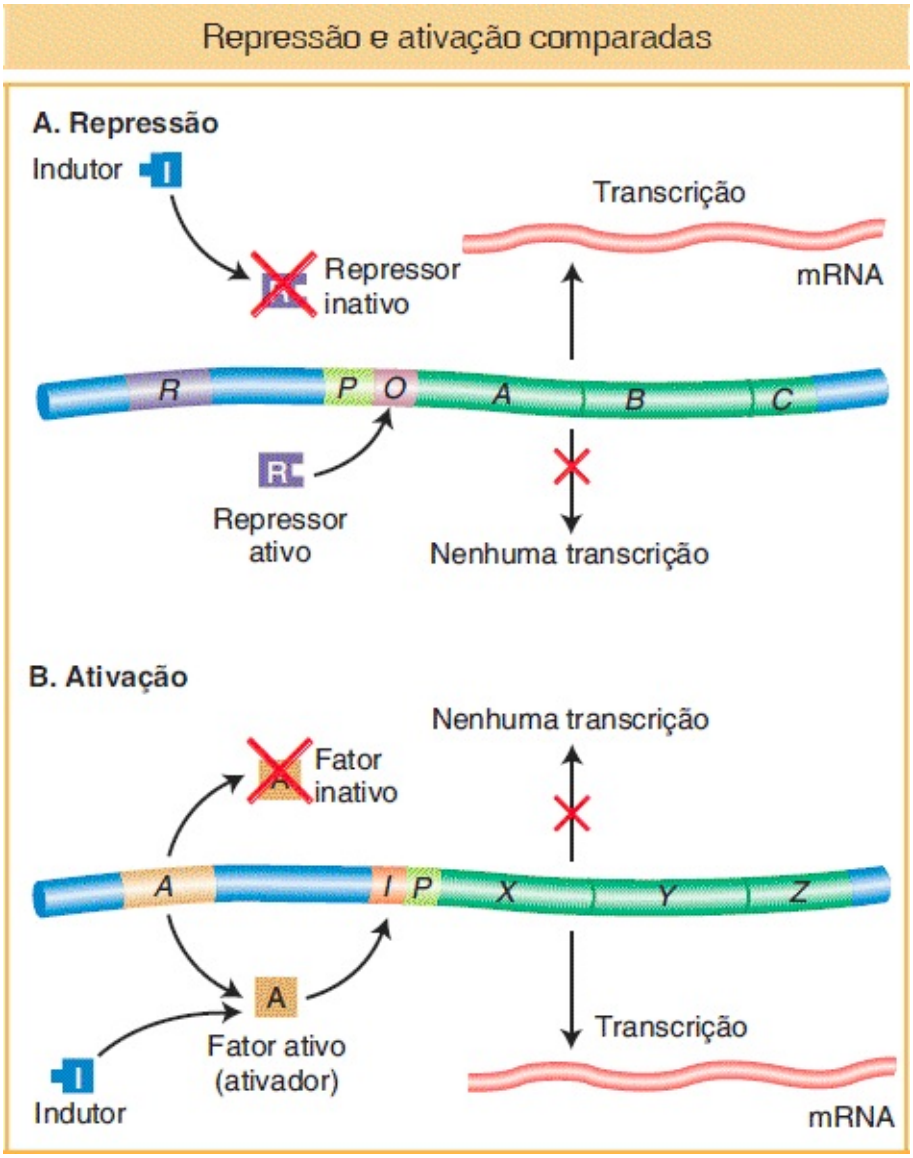


FIGURA 11.18 A. Na repressão, um repressor ativo (codificado pelo gene *R* neste exemplo) bloqueia a expressão do óperon *A*, *B*, *C* por meio da ligação a um sítio operador (*O*). **B.** Na ativação, é necessário um ativador funcional para a expressão gênica. Um ativador não funcional resulta em ausência de expressão dos genes *X*, *Y*, *Z*. Pequenas moléculas conseguem converter um ativador não funcional em um ativador funcional, que em seguida se liga à região de controle do óperon, neste caso denominada *I*. As posições tanto de *O* quanto de *I* em relação ao promotor *P* nos dois exemplos são desenhadas arbitrariamente, visto que as suas posições diferem em diferentes óperons.

11.4 Duplo controle positivo e negativo | Óperon da arabinose

Assim como com o sistema *lac*, o controle da transcrição em bactérias não é nem puramente positivo, nem puramente negativo; em vez disso, tanto a regulação positiva quanto a negativa podem regular óperons individuais. A regulação do óperon da arabinose fornece um exemplo no qual uma única proteína de ligação ao DNA pode atuar *tanto* como um repressor *quanto* como um ativador — um desvio no tema geral da regulação da transcrição por proteínas de ligação ao DNA.

Os genes estruturais *araB*, *araA* e *araD* codificam as enzimas metabólicas que fragmentam o açúcar arabinose. Os três genes são transcritos em uma unidade como um único mRNA. A Figura 11.19 demonstra um mapa do óperon *ara*. A transcrição é ativada em *araI*, a região **iniciadora**, que contém um sítio de ligação para uma proteína ativadora. O gene *araC*, que está mapeado próximo, codifica uma proteína ativadora. Quando ligada à arabinose, essa proteína se liga ao sítio *araI* e ativa a transcrição do óperon *ara*, talvez por auxiliar a RNA polimerase a se ligar ao promotor. Além disso, o mesmo sistema de repressão catabólica CAP-cAMP que evita a expressão do óperon *lac* na presença de glicose também evita a expressão do óperon *ara*.

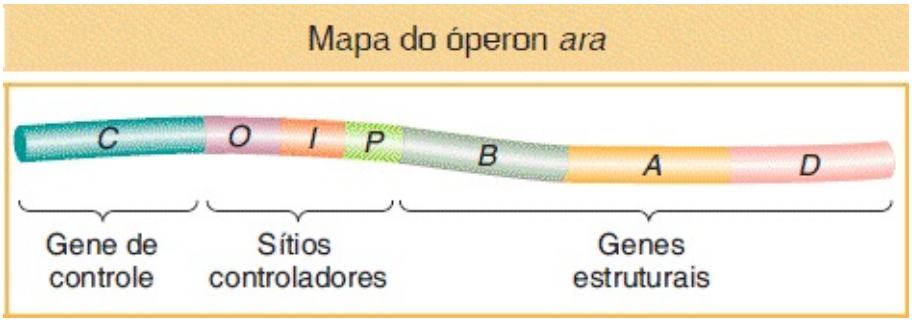


FIGURA 11.19 Os genes *B*, *A* e *D* juntamente com os sítios *I* e *O* constituem o óperon *ara*. *O* é *araO* e *I* é *araI*.

Na presença de arabinose, ambos os complexos CAP-cAMP e AraC-arabinose devem se ligar a *araI* para que a RNA polimerase se ligue ao promotor e transcreva o óperon *ara* (Figura 11.20 A). Na ausência de arabinose, a proteína AraC assume uma conformação diferente e reprime o óperon *ara* por meio da ligação a *araI* e a um segundo sítio distante, *araO*, formando, assim, uma alça

(Figura 11.20 B) que evita a transcrição. Portanto, a proteína AraC apresenta duas conformações, uma que atua como um ativador e outra que atua como um repressor. A alternância liga/desliga do óperon é "disparada" pela arabinose. As duas conformações, dependendo de a arabinose efetora alostérica estar ligada à proteína, diferem em suas capacidades de se ligar a um sítio-alvo específico na região *araO* do óperon.

> CONCEITO-CHAVE A transcrição do óperon pode ser regulada por ambas a ativação e a repressão. Os óperons que regulam o metabolismo de compostos semelhantes, tais como açúcares, podem ser regulados de modos consideravelmente diferentes.

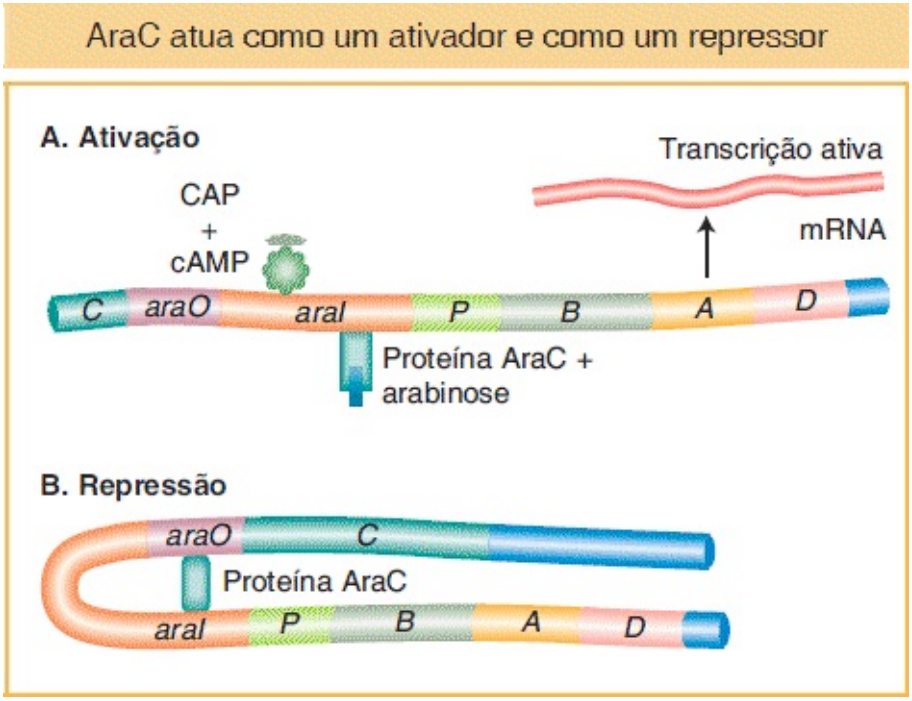


FIGURA 11.20 Duplo controle do óperon *ara*. **A.** Na presença de arabinose, a proteína AraC se liga à região *araI*. O complexo CAP-cAMP se liga a um sítio adjacente a *araI*. Esta ligação estimula a transcrição dos genes *araB*, *araA* e *araD*. **B.** Na ausência de arabinose, a proteína AraC se liga a ambas as regiões *araI* e *araO*, formando uma alça de DNA. Esta ligação evita a transcrição do óperon *ara*.

Vias metabólicas e níveis adicionais de regulação | 11.5 Atenuação

O controle coordenado dos genes em bactérias é difuso. Na seção precedente, observamos exemplos que ilustram a regulação de vias para a fragmentação de açúcares específicos. De fato, a maior parte dos genes coordenados em bactérias é regulada por meio de mecanismos de óperon. Em muitas vias que sintetizam moléculas essenciais de blocos de construção inorgânicos simples, os genes que codificam as enzimas estão organizados em óperons, completados com mRNA multigênicos. Além disso, nos casos em relação aos quais a sequência da atividade catalítica é conhecida, existe uma extraordinária congruência entre a ordem dos genes no óperon no cromossomo e a ordem na qual os seus produtos atuam na via metabólica. Essa congruência está surpreendentemente ilustrada pela organização do óperon do triptofano em *E. coli* (Figura 11.21). O óperon do triptofano contém cinco genes (*trpE, trpD, trpC, trpB, trpA*) que codificam enzimas que contribuem para a síntese do aminoácido triptofano.

> CONCEITO-CHAVE Em bactérias, os genes que codificam enzimas que estão nas mesmas vias metabólicas em geral estão organizados em óperons.

Existem dois mecanismos para regular a transcrição no óperon do triptofano e em alguns outros óperons que atuam na biossíntese de aminoácidos. Um proporciona o controle global da expressão do mRNA do óperon, e o outro proporciona um controle fino.

O nível de expressão gênica do óperon *trp* é regulado pelo nível de triptofano. Quando o triptofano está ausente no meio de crescimento, a expressão do gene *trp* é alta; quando os níveis de triptofano são altos, o óperon *trp* é reprimido. Um mecanismo para o controle da transcrição do óperon *trp* é semelhante ao mecanismo que já vimos e que controla o óperon *lac*: uma proteína repressora se liga a um operador, evitando o início da transcrição. Esse repressor é o repressor

Trp, o produto do gene *trpR*. O repressor Trp se liga ao triptofano quando estão presentes níveis adequados do aminoácido, e apenas após a ligação do triptofano o repressor Trp se ligará ao operador e desligará a transcrição do óperon. Esse mecanismo simples assegura que a célula não desperdice energia produzindo triptofano quando o aminoácido é suficientemente abundante. Linhagens de *E. coli* com mutações em *trpR* continuam a expressar o mRNA de *trp* e, portanto, continuam a produzir triptofano quando o aminoácido é abundante.

Ao estudar essas linhagens mutantes de *trpR*, Charles Yanofsky descobriu que, quando o triptofano era removido do meio, a produção de mRNA de *trp* aumentava várias vezes. Esse achado foi a evidência de que, além do repressor Trp, existia um segundo mecanismo de controle para regular negativamente a transcrição. Esse mecanismo é denominado **atenuação**, tendo em vista que a produção de mRNA normalmente é *atenuada*, o que significa "diminuída", quando o triptofano é abundante. Contrariamente aos outros mecanismos de controle bacterianos descritos até agora, a atenuação atua como uma etapa *após* o início da transcrição.

Os mecanismos que regulam a atenuação foram descobertos por meio da identificação de mutações que reduziam ou aboliam a atenuação. Linhagens com essas mutações produzem mRNA de *trp* em níveis máximos mesmo na presença de triptofano. Yanofsky mapeou as mutações em uma região entre o operador *trp* e o gene *trpE*; essa região, denominada **sequência líder**, está na extremidade 5′ do mRNA do óperon *trp*, antes do primeiro códon do gene *trpE* (Figura 11.22). A sequência líder de *trp* é incomumente longa para um mRNA procariótico, de 160 bases, e análises detalhadas revelaram como uma parte dessa sequência atua como um **atenuador** que regula a transcrição adicional do mRNA de *trp*.

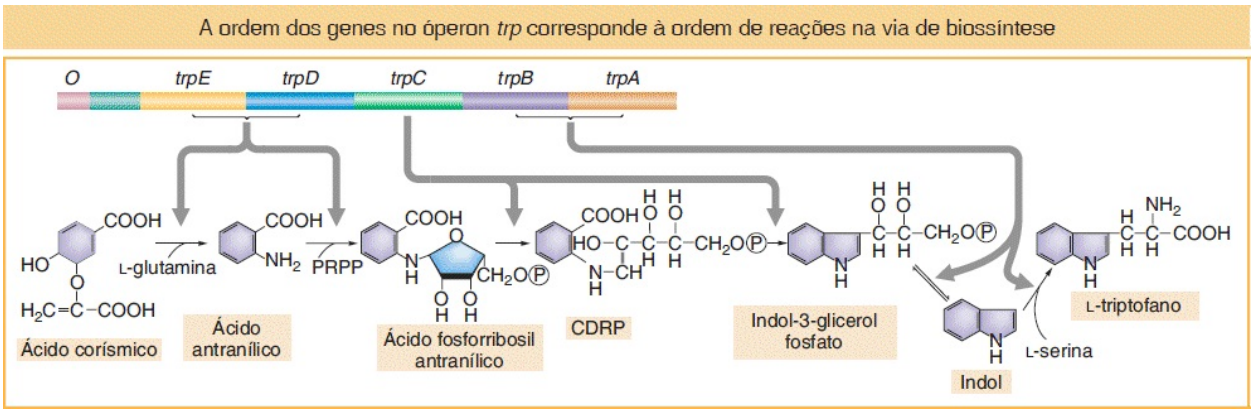


FIGURA 11.21 A ordem cromossômica dos genes no óperon *trp* de *E. coli* e a sequência de reações catalisadas pelos produtos enzimáticos dos genes estruturais *trp*. Os produtos dos genes *trpD* e *trpE* formam um complexo que catalisa etapas específicas, assim como os produtos dos genes *trpB* e *trpA*. A triptofano sintetase é uma enzima tetramérica formada pelos produtos de *trpB* e *trpA*. Ela catalisa um processo de duas etapas que leva à formação do triptofano. Abreviações: PRPP = fosforribosilpirofosfato; CDRP = 1-(*o*-carboxifenilamino)-1-desoxirribulose 5-fosfato. (Dados de S. Tanemura e R. H. Bauerle, *Genetics* 95, 1980, 545.)

As observações-chave são que, na ausência da proteína repressora TrpR, a presença do triptofano interrompe a transcrição após aproximadamente as primeiras 140 bases, enquanto, na ausência de triptofano, a transcrição do óperon continua. O mecanismo de término ou de continuação da transcrição é composto por dois elementos-chave. Primeiramente, a **sequência líder** do mRNA de *trp* codifica um peptídio curto de 14 aminoácidos, que inclui dois códons de triptofano adjacentes. O triptofano é um dos aminoácidos menos abundantes nas proteínas, e é codificado por um único códon. Esse par de códons de triptofano, portanto, é uma característica incomum. Em segundo lugar, partes do mRNA líder *trp* formam estruturas de haste e alça que são capazes de alternar entre duas conformações. Uma dessas conformações favorece o término da transcrição (Figura 11.23 A).

A lógica reguladora do óperon depende da abundância de triptofano. Quando o triptofano é abundante, existe um suprimento suficiente de aminoacil-tRNA^{Trp} para possibilitar a tradução do peptídio de 14 aminoácidos. Relembre que a transcrição e a tradução em bactérias estão acopladas; assim, os ribossomos podem ligar-se aos transcritos de mRNA e iniciar a tradução antes que a

transcrição esteja completa. A ligação do ribossomo altera a conformação do mRNA de *trp* de tal forma que favorece o término da transcrição (onde os segmentos 3 e 4 formam pares de bases; Figura 11.23 B). Entretanto, quando o triptofano é escasso, o ribossomo fica parado nos códons do triptofano, os segmentos 2 e 3 pareiam bases, e a transcrição pode continuar (Figura 11.23 C).

Outros óperons para enzimas nas vias de biossíntese apresentam controles de atenuação semelhantes. Uma característica dos óperons de biossíntese de aminoácidos é a presença de múltiplos códons para o aminoácido que está sendo sintetizado em um peptídio separado codificado pela sequência líder 5′. Por exemplo, o óperon *phe* apresenta sete códons de fenilalanina em um peptídio líder e o óperon *his* apresenta sete códons de histidina em *tandem* em seu peptídio líder (Figura 11.24).

> CONCEITO-CHAVE Um segundo nível de regulação em óperons de biossíntese de aminoácidos é a atenuação da transcrição mediada pela abundância do aminoácido e pela tradução de um peptídio líder.

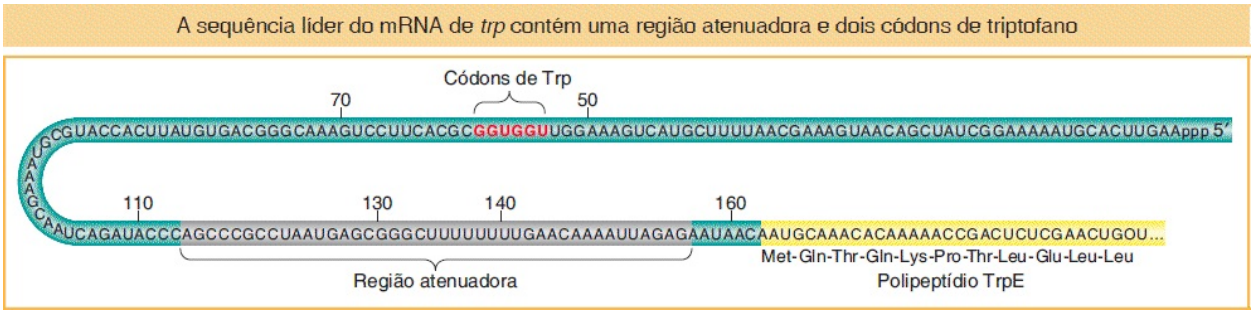


FIGURA 11.22 Na sequência líder do mRNA de *trp*, a região atenuadora precede a sequência codificadora *trpE*. *Upstream*, nas bases 54 a 59, estão os dois códons de triptofano (*demonstrados em vermelho*) do peptídio líder.

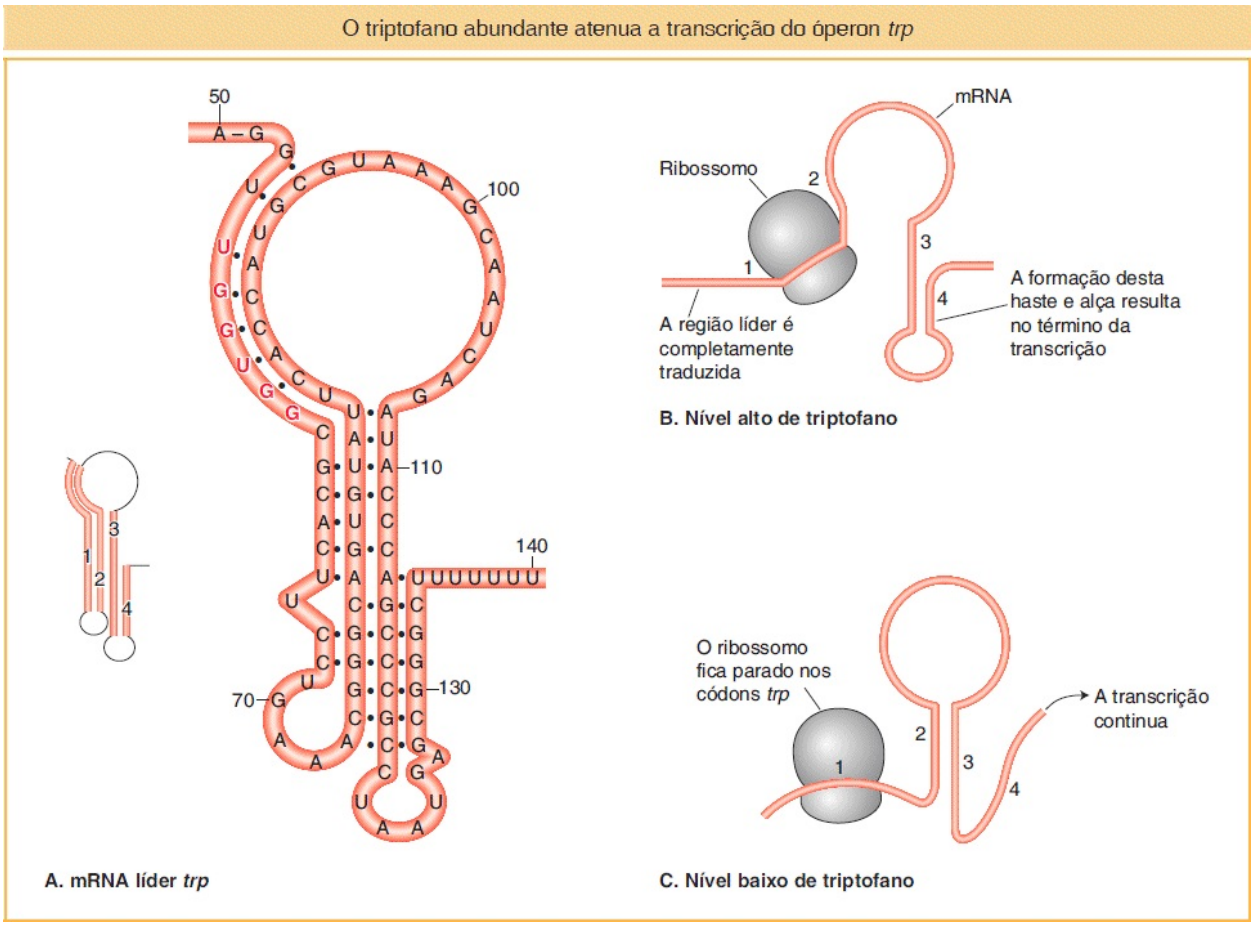


FIGURA 11.23 Modelo em relação à atenuação no óperon *trp*. **A.** Estruturas secundárias propostas na conformação do mRNA líder do *trp* que favorecem o término da transcrição. Quatro regiões podem parear bases para formar três estruturas de haste e alça, mas apenas duas regiões pareiam bases entre si em um determinado momento. Portanto, a região 2 pode parear bases com a região 1 ou a região 3. **B.** Quando o triptofano é abundante, o segmento 1 do mRNA do *trp* é traduzido. O segmento 2 entra no ribossomo (embora não seja traduzido), o que possibilita que os segmentos 3 e 4 pareiem bases. Esta região com bases pareadas faz com que a RNA polimerase termine a transcrição. **C.** Contrariamente, quando o triptofano é escasso, o ribossomo fica parado nos códons do segmento 1. O segmento 2 interage com o segmento 3 em vez de ser puxado para o ribossomo e, assim, os segmentos 3 e 4 não podem parear. Consequentemente, a transcrição continua. (Dados de D. L. Oxender, G. Zurawsk, e C. Yanofsky, *Proc. Natl. Acad. Sci.* USA 76, 1979, 5524.)

FIGURA 11.24 A. A parte traduzida da região líder *trp* contém dois códons de triptofano consecutivos. **B.** A sequência líder *phe* contém sete códons de fenilalanina. **C.** A sequência líder *his* contém sete códons de histidina consecutivos.

11.6 Ciclos de vida de bacteriófagos | Mais reguladores, óperons complexos

Naquele cinema em Paris, François Jacob teve um lampejo de que o fenômeno da indução do prófago pudesse ser análogo de modo próximo à indução da síntese de β-galactosidase. Ele estava certo. Aqui, veremos como o ciclo de vida do bacteriófago λ é regulado. Embora a sua regulação seja mais complexa do que aquela dos óperons individuais, ela é controlada por modos de regulação gênica agora familiares.

O bacteriófago λ é um assim denominado fago temperado, que apresenta dois ciclos de vida alternativos (Figura 11.25). Quando uma bactéria normal é infectada por um fago λ do tipo selvagem, podem se seguir dois possíveis desfechos: (1) o fago pode se replicar e finalmente lisar a célula (o **ciclo lítico**), ou (2) o genoma do fago pode integrar-se ao cromossomo bacteriano como um prófago inerte (o **ciclo lisogênico**). No estado lítico, a maior parte dos 71 genes do fago é expressa em algum ponto, enquanto no estado lisogênico, a maior parte dos genes está inativa.

O que decide qual dessas duas vias é adotada? O controle fisiológico da decisão entre a via lítica ou lisogênica depende dos recursos disponíveis na

bactéria hospedeira. Se os recursos forem abundantes, o ciclo lítico é preferido, tendo em vista que então existem nutrientes suficientes para produzir muitas cópias do vírus. Se os recursos forem limitados, a via lisogênica é adotada. O vírus então permanece presente como um *prófago* até que as condições melhorem. O prófago inerte pode ser induzido por luz ultravioleta a entrar no ciclo lítico — o fenômeno estudado por Jacob. Os estados lítico e lisogênico são caracterizados por programas muito distintos de expressão gênica que devem ser regulados. Qual estado alternativo é selecionado é determinado por um interruptor genético complexo que compreende diversas proteínas reguladoras de ligação ao DNA e um conjunto de sítios operadores.

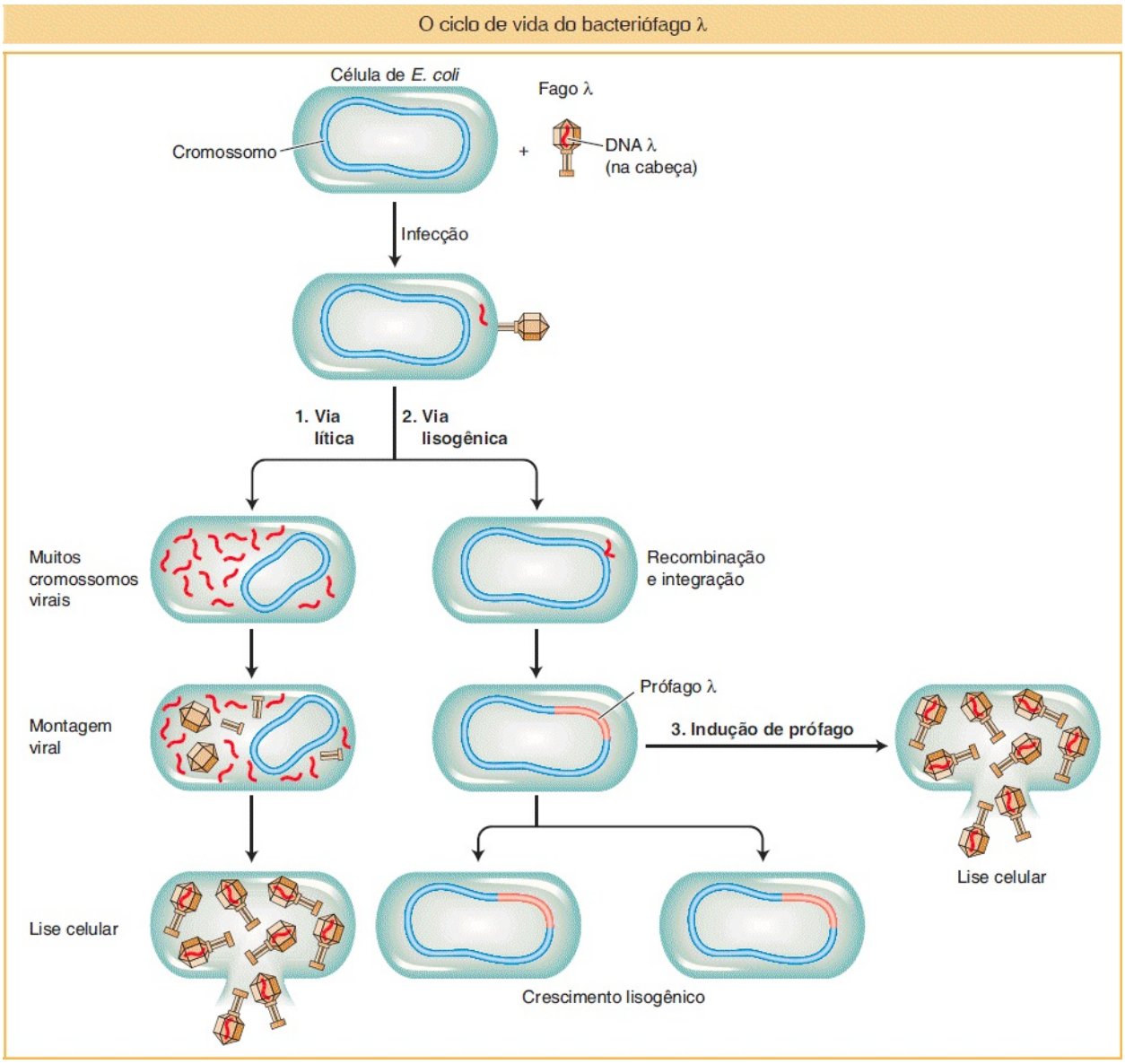


FIGURA 11.25 Se o bacteriófago λ entra imediatamente no ciclo lítico ou se entra no ciclo lisogênico depende da disponibilidade de recursos. O vírus lisogênico insere seu genoma no cromossomo bacteriano, onde ele permanece quiescente até que as condições sejam favoráveis.

Assim como o foram para os sistemas reguladores *lac* e outros, as análises genéticas de mutantes foram fonte de percepções cruciais sobre os componentes e a lógica do interruptor genético de λ. Jacob utilizou triagens fenotípicas simples para isolar mutantes que eram defeituosos na via lítica ou na lisogênica. Os mutantes de cada tipo puderam ser reconhecidos pelo aspecto das placas infectadas em uma camada de bactérias. Quando partículas de fago do tipo

selvagem são colocadas em uma camada de bactérias sensíveis, aparecem áreas claras (denominadas “placas”) onde as bactérias são infectadas e lisadas, mas essas placas são turvas, tendo em vista que bactérias lisogenizadas crescem nelas (Figura 11.26). Os fagos mutantes que formam placas claras são incapazes de lisogenizar as células.

Os referidos mutantes *claros* (denominados por *c*) são análogos aos mutantes *I* e *O* do sistema *lac*. Esses mutantes com frequência foram isolados como mutantes sensíveis à temperatura que apresentavam fenótipos *claros* em temperaturas mais altas, mas fenótipos do tipo selvagem em temperaturas mais baixas. Três classes de mutantes levaram à identificação das características reguladoras-chave do fago λ. Na primeira classe, os mutantes para os genes *cI*, *cII* e *cIII* formam placas claras; ou seja, eles são incapazes de estabelecer a lisogenia. Foi isolada uma segunda classe de mutantes que não lisogeniza as células, mas que pode se replicar e entrar no ciclo lítico em uma célula lisogenizada. Esses mutantes são análogos aos mutantes constitutivos do operador do sistema *lac*. Um terceiro mutante-chave consegue lisogenizar, mas é incapaz de lisar as células. O gene mutado nesse caso é o gene *cro* (em referência ao *c*ontrole do *r*epressor e de *o*utras coisas). A decisão entre as vias lítica e lisogênica depende da atividade das proteínas codificadas pelos quatro genes *cI*, *cII*, *cIII*, e *cro*, três dos quais são proteínas de ligação ao DNA.

FIGURA 11.26 As placas são claras onde ocorreu a lise da célula hospedeira; elas são turvas onde as células sobreviveram à infecção e continuaram a crescer como um lisógeno. (De *Microbiology An Evolving Science*, 1st ed., Figure 10.22. © John Foster.)

Primeiramente enfocaremos os dois genes *cI* e *cro* nas proteínas que eles codificam (Tabela 11.4). O gene *cI* codifica um repressor, com frequência denominado repressor λ, que reprime o crescimento lítico e promove a lisogenia. O gene *cro* codifica um repressor que reprime a lisogenia, possibilitando, assim, o crescimento lítico. O interruptor genético que controla os dois ciclos de vida do fago λ apresenta dois estados: no estado lisogênico, *cI* está ligado, mas *cro* está desligado, e no ciclo lítico, *cro* está ligado, mas *cI* está desligado. Portanto, o repressor λ e Cro estão competindo, e o repressor que prevalece determina o estado do interruptor e da expressão do genoma λ.

A corrida entre o repressor λ e o Cro é iniciada quando o fago λ infecta uma bactéria normal. A sequência de eventos na corrida é criticamente determinada pela organização dos genes no genoma λ e dos promotores e operadores entre os genes *cI* e *cro*. O genoma λ de aproximadamente 50 kb codifica proteínas que apresentam papéis na replicação do DNA, na recombinação, na montagem da

partícula do fago e na lise celular (Figura 11.27). Essas proteínas são expressas em uma sequência lógica tal que primeiramente são produzidas cópias do genoma; essas cópias em seguida são empacotadas em partículas virais e, finalmente, a célula hospedeira é lisada para liberar o vírus e iniciar a infecção de outras células hospedeiras (ver Figura 11.25). A ordem da expressão gênica viral flui a partir do início da transcrição em dois promotores, P_L e P_R (em referência a promotor à esquerda e à direita em relação ao mapa genético). Com a infecção, a RNA polimerase inicia a transcrição em ambos os promotores. Observando o mapa genético (ver Figura 11.27), vemos que a partir de P_R, *cro* é o primeiro gene transcrito, e a partir de P_L, *N* é o primeiro gene transcrito.

O gene *N* codifica um regulador positivo, mas o mecanismo dessa proteína difere daquele de outros reguladores que consideramos até agora. A proteína N possibilita que a RNA polimerase continue a transcrever regiões do DNA que de outro modo causariam o término da transcrição. Uma proteína reguladora tal como N, que atua impedindo o término da transcrição, é denominada **antiterminalizador**. Portanto, *N* possibilita a transcrição de *cIII* e outros genes à esquerda de *N*, bem como de *cII* e outros genes à direita de *cro*. O gene *cII* codifica uma proteína ativadora que se liga a um sítio que promove a transcrição à esquerda de um promotor diferente, P_{RE} (em referência a promotor de *e*stabelecimento do *r*epressor), que ativa a transcrição do gene *cI*. Relembre que o gene *cI* codifica o repressor λ, que evitará o crescimento lítico.

Tabela 11.4 Principais reguladores do ciclo de vida do bacteriófago λ.

Gene	Proteína	Promove
cI	repressor λ	via lisogênica
cro	repressor Cro	via lítica
N	regulador positivo	expressão de cII, cIII

cII	ativador	expressão de cI
cIII	inibidor de protease	atividade de cII

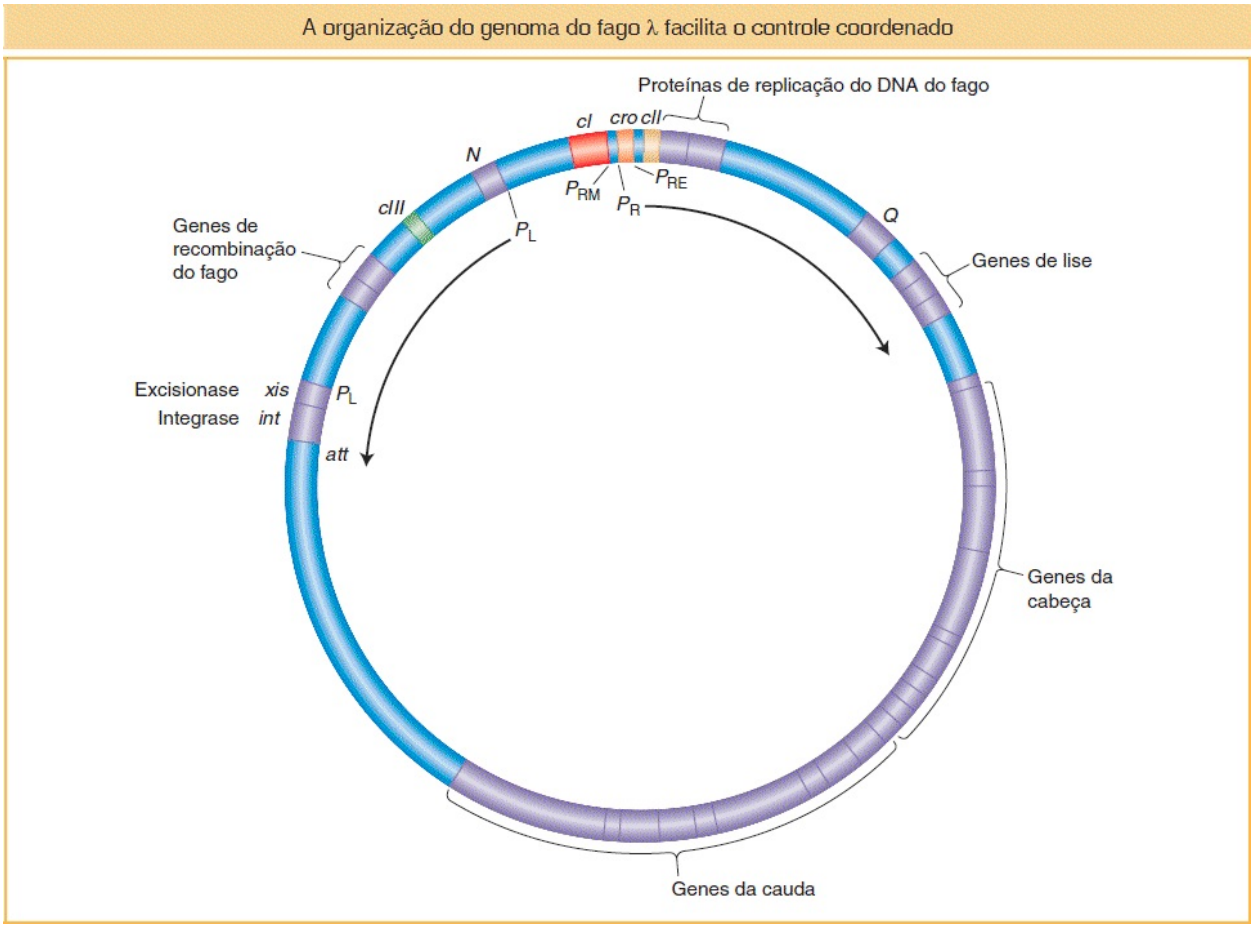


FIGURA 11.27 Mapa do fago λ na forma circular. Os genes para recombinação, integração e excisão, replicação, montagem da cabeça e da cauda e lise celular estão agrupados juntos e são regulados de modo coordenado. A transcrição do lado direito do genoma tem início em P_R, e aquela dos genes à esquerda tem início em P_L. Interações reguladoras-chave que regulam a decisão lisogênica *versus* lítica ocorrem nos operadores entre os genes *cro* e *cI*.

Antes de ocorrer a expressão do restante dos genes virais, deve ser tomada uma "decisão" — continuar com a expressão do gene viral e lisar a célula, ou reprimir aquela via e lisogenizar a célula. A decisão de lisar ou lisogenizar uma

célula depende da atividade da proteína cII. A proteína cII é instável, tendo em vista que é sensível a proteases bacterianas — enzimas que degradam proteínas. Essas proteases respondem às condições ambientais: elas são mais ativas quando os recursos são abundantes, mas menos ativas quando as células estão em inanição.

Vejamos o que ocorre com cII quando os recursos são abundantes e não abundantes. Quando os recursos são abundantes, cII é degradado e pouco repressor λ é produzido. Os genes transcritos a partir de P_L e P_R continuam a ser expressos, e o ciclo lítico prevalece. Entretanto, se os recursos são limitados, cII é mais ativo e mais repressor λ é produzido. Nesse caso, os genes transcritos a partir de P_L e P_R são reprimidos pelo repressor λ e ocorre a entrada no ciclo lisogênico. A proteína cII também é responsável pela ativação da transcrição de *int*, um gene que codifica uma proteína adicional necessária para a lisogenia — uma integrase necessária para a integração do genoma λ ao cromossomo hospedeiro. A proteína cIII protege cII da degradação; assim, ela também contribui para a decisão lisogênica.

Recapitulemos brevemente a sequência de eventos e os pontos de decisão no ciclo de vida do bacteriófago λ:

Com a infecção:

A RNA polimerase do hospedeiro inicia a transcrição em P_L e P_R, expressando os genes *cro* e *N*.

Em seguida:

A proteína antiterminalizadora N possibilita a transcrição do gene *cIII* e dos genes de recombinação (ver Figura 11.27, *esquerda*), bem como do gene *cII* e outros genes (ver Figura 11.27, *direita*).

Em seguida:

A proteína cII, protegida pela proteína cIII, liga *cI* e *int* por meio da ativação da transcrição em P_{RE}.

Em seguida:

Se os recursos e as proteases forem abundantes, cII é degradada, Cro reprime *cI*, e o ciclo lítico continua.

Se os recursos e as proteases não forem abundantes, cII está ativa, a transcrição de *cI* prossegue em um nível alto, a proteína Int integra-se ao cromossomo do fago, e cI (repressor λ) desliga todos os genes, exceto a si próprio.

Anatomia molecular do interruptor genético

Para ver como a decisão é executada no nível molecular, nos voltaremos para as atividades do repressor λ e de Cro. O operador O_R está localizado entre os dois genes que codificam essas proteínas e contém três sítios, O_{R1}, O_{R2} e O_{R3}, que sobrepõem dois promotores opostos: P_R, que promove a transcrição de genes líticos, e P_{RM} (em referência a *m*anutenção de *r*epressor), que direciona a transcrição do gene *cI* (ver Figura 11.27). Relembre que o gene *cI* codifica o repressor λ. Os três sítios operadores são semelhantes, mas não idênticos em sequência, e embora Cro e o repressor λ possam se ligar a qualquer um dos operadores, eles realizam isso com diferentes afinidades: o repressor λ se liga a O_{R1} com a mais alta afinidade, enquanto Cro se liga a O_{R3} com a mais alta afinidade. A ocupação de O_{R1} pelo repressor λ bloqueia a transcrição a partir de P_R e, portanto, bloqueia a transcrição dos genes para o sítio lítico. A ocupação de O_{R3} por Cro bloqueia a transcrição a partir de P_{RM} e, portanto, bloqueia a manutenção da transcrição de *cI*. Portanto, nenhum repressor λ é produzido, e a transcrição dos genes em relação ao ciclo lítico pode continuar. A ocupação dos sítios operadores, portanto, determina os padrões lítico *versus* lisogênico da expressão do gene λ (Figura 11.28).

Após um lisógeno ter sido estabelecido, em geral ele é estável. Mas o lisógeno pode ser induzido a entrar no ciclo lítico por meio de diversas alterações ambientais. A luz ultravioleta induz a expressão dos genes do hospedeiro. Um dos genes do hospedeiro codifica uma proteína, RecA, que estimula a clivagem do repressor λ, afetando, assim, a manutenção da lisogenia e resultando em crescimento lítico. A indução do prófago, assim como Jacob e Monod conjeturaram, requer a liberação de um repressor do DNA. O papel fisiológico da luz ultravioleta na indução lisogênica faz sentido, pois esse tipo de radiação lesiona o DNA hospedeiro e estressa as bactérias; o fago replica e deixa a célula lesionada e estressada para outro hospedeiro.

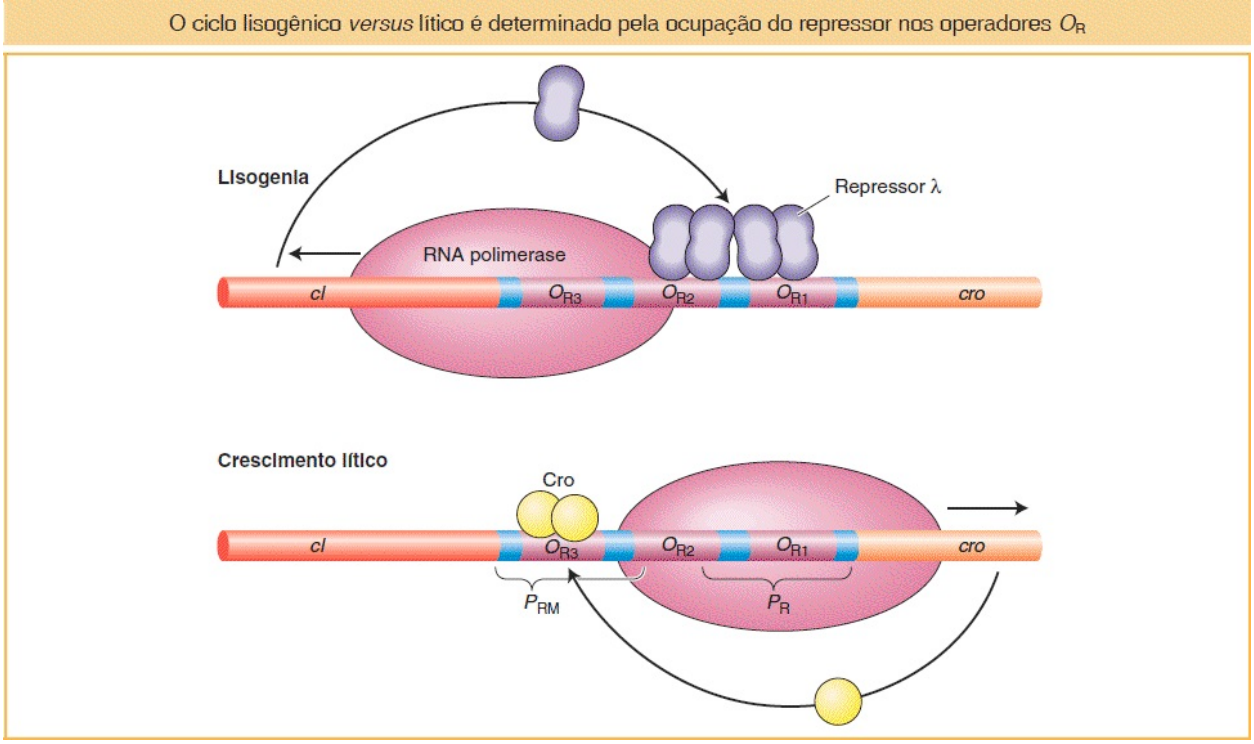


FIGURA 11.28 A ligação do repressor λ e de Cro aos sítios operadores. A lisogenia é promovida pela ligação do repressor λ a O_{R1} e O_{R2}, o que evita a transcrição a partir de P_R. Na indução ou no ciclo lítico, a ligação de Cro a O_{R3} evita a transcrição do gene *cI*. (Dados de M. Ptashne e A. Gann, *Genes and Signals*, p. 30, Fig. 1-13.)

CONCEITO-CHAVE O interruptor genético do fago λ ilustra como algumas poucas proteínas reguladoras de ligação ao DNA, que atuam por meio de alguns poucos sítios, controlam a expressão de um número muito maior de genes no vírus, em um mecanismo de "cascata". Assim como nos sistemas *lac*, *ara*, *trp* e outros, os estados alternativos da expressão gênica são determinados por sinais fisiológicos.

Ligação sequência-específica de proteínas reguladoras ao DNA

Como o repressor λ e Cro reconhecem diferentes operadores com diferentes afinidades? Essa questão direciona a nossa atenção para um princípio fundamental no controle da transcrição gênica — as proteínas reguladoras se ligam a sequências de DNA específicas. Para que proteínas individuais se liguem

a determinadas sequências e não a outras, há necessidade de especificidade nas interações das cadeias laterais dos aminoácidos da proteína e os grupos químicos das bases do DNA. Estudos estruturais detalhados do repressor λ, Cro e de outros reguladores bacterianos revelaram como as estruturas tridimensionais dos reguladores e do DNA interagem e como o arranjo de aminoácidos em particular lhes possibilita reconhecer sequências de bases específicas.

A análise cristalográfica identificou uma característica estrutural comum dos domínios de ligação ao DNA de λ e Cro. Ambas as proteínas fazem contato com o DNA por meio de um domínio de *hélice-giro-hélice*, que é composto por duas hélices α unidas por uma região ligante flexível curta (Figura 11.29). Uma hélice, a hélice de reconhecimento, se encaixa no sulco maior do DNA. Naquela posição, os aminoácidos na face externa da hélice são capazes de interagir com grupos químicos nas bases do DNA. Os aminoácidos específicos na hélice de reconhecimento determinam a afinidade de uma proteína por uma sequência de DNA específica.

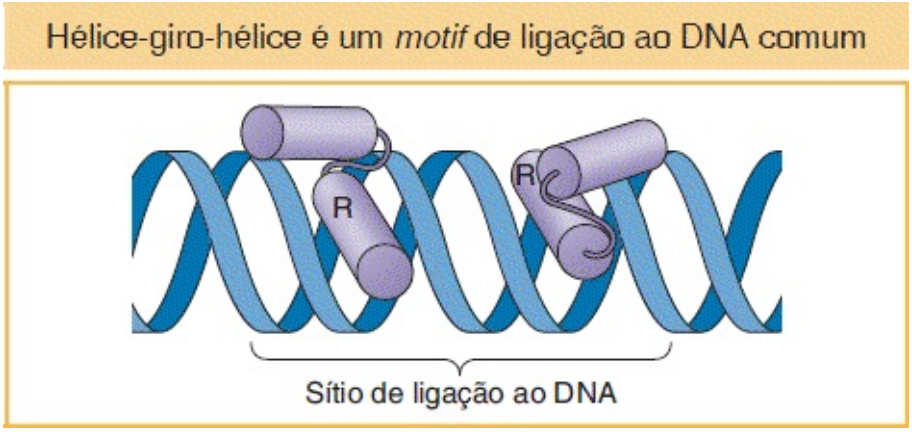


FIGURA 11.29 A ligação de um *motif* hélice-giro-hélice ao DNA. Os *cilindros roxos* são as hélices alfa. Muitas proteínas reguladoras se ligam como dímeros ao DNA. Em cada monômero, a hélice de reconhecimento (R) faz contato com bases no sulco maior do DNA.

As hélices de reconhecimento do repressor λ e do Cro apresentam estruturas semelhantes e alguns resíduos de aminoácidos idênticos. As diferenças entre as hélices nos resíduos de aminoácidos-chave determinam as suas propriedades de ligação ao DNA. Por exemplo, no repressor λ e proteínas Cro, as cadeias laterais

de glutamina e serina fazem contato com as mesmas bases, mas um resíduo de alanina no repressor λ e resíduos de lisina e asparagina na proteína Cro impõem afinidades de ligação diferentes para sequências em O_{R1} e O_{R3} (Figura 11.30).

Os repressores Lac e TrpR, bem como o ativador AraC e muitas outras proteínas, também se ligam ao DNA por meio de *motifs* hélice-giro-hélice de diferentes especificidades, dependendo das sequências de aminoácidos primários de suas hélices de reconhecimento. Em geral, outros domínios dessas proteínas, tais como aqueles que se ligam a seus respectivos efetores alostéricos, são diferentes.

> CONCEITO-CHAVE A especificidade biológica da regulação gênica ocorre em virtude da especificidade química das interações aminoácido-base de proteínas reguladoras individuais com as sequências de DNA discretas.

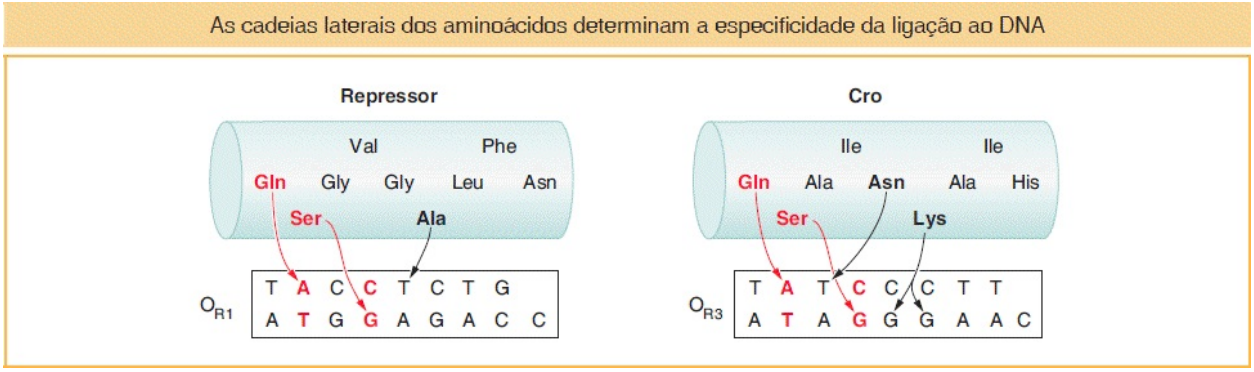


FIGURA 11.30 As interações de aminoácidos e bases determinam a especificidade e a afinidade das proteínas de ligação ao DNA. Estão demonstradas as sequências de aminoácidos das hélices de reconhecimento do repressor λ e das proteínas Cro. As interações dos resíduos de glutamina (Gln), serina (Ser) e alanina (Ala) do repressor λ e as bases no operador O_R determinam a força da ligação. De modo semelhante, as interações dos resíduos de glutamina, serina, asparagina (Asn) e lisina (Lys) da proteína Cro medeiam a ligação ao operador O_{R3}. Cada sequência de DNA demonstrada é aquela ligada por um monômero individual do respectivo repressor; ela tem metade do sítio operador ocupado pelo dímero do repressor. (Dados de M. Ptashne, *A Genetic Switch: Phage I and Higher Organisms*, 2nd ed.)

11.7 Fatores sigma alternativos regulam grandes

conjuntos de genes

Até agora, vimos como interruptores únicos podem controlar a expressão de óperons únicos ou de dois óperons que contêm tantos quanto um par de dúzias de genes. Algumas respostas fisiológicas às alterações no ambiente requerem a expressão coordenada de grandes conjuntos de genes não ligados localizados por todo o genoma para ocasionar alterações fisiológicas dramáticas e até mesmo morfológicas. As análises desses processos revelaram outro desvio na regulação gênica bacteriana: o controle de grandes números de genes por fatores sigma (σ) alternativos da RNA polimerase. Um referido exemplo, o processo da esporulação no *Bacillus subtilis*, foi analisado em grande detalhe nas últimas poucas décadas. Sob estresse, a bactéria forma esporos que são notavelmente resistentes ao calor e à dessecação.

Inicialmente no processo de esporulação, a bactéria se divide assimetricamente, gerando dois componentes de tamanho desigual que apresentam destinos muito diferentes. O compartimento menor, o pró-esporo, desenvolve-se dentro do esporo. O compartimento maior, a célula-mãe, nutre o esporo em desenvolvimento e é lisado quando a morfogênese do esporo está completa, para liberar o esporo (Figura 11.31 A). A dissecção genética desse processo envolveu o isolamento de muitos mutantes que não conseguem esporular. Investigações detalhadas levaram à caracterização de diversas proteínas reguladoras-chave que regulam diretamente programas de expressão gênica que são específicos do pré-esporo ou da célula-mãe. Quatro dessas proteínas são fatores σ alternativos.

Lembre que o início da transcrição em bactérias inclui a ligação da subunidade σ da RNA polimerase às regiões —35 e —10 dos promotores dos genes. O fator σ se desassocia do complexo quando a transcrição tem início e é reciclado. No *B. subtilis*, dois fatores σ, σ^A e σ^H, são ativos em células vegetativas. Durante a esporulação, um diferente fator σ, σ^F, torna-se ativo no pró-esporo e ativa um grupo de mais de 40 genes. Um gene ativado por σ^F é uma proteína secretada que, por sua vez, aciona o processamento proteolítico do precursor inativo pró-σ^E, um fator σ distinto na célula-mãe. O fator σ^E é necessário para ativar conjuntos de genes na célula-mãe. Dois fatores σ adicionais, σ^K e σ^G, são subsequentemente

ativados na célula-mãe e no pró-esporo, respectivamente (Figura 11.31 A). A expressão de fatores σ distintos possibilita a transcrição coordenada de diferentes conjuntos de genes, ou **regulons**, por uma única RNA polimerase.

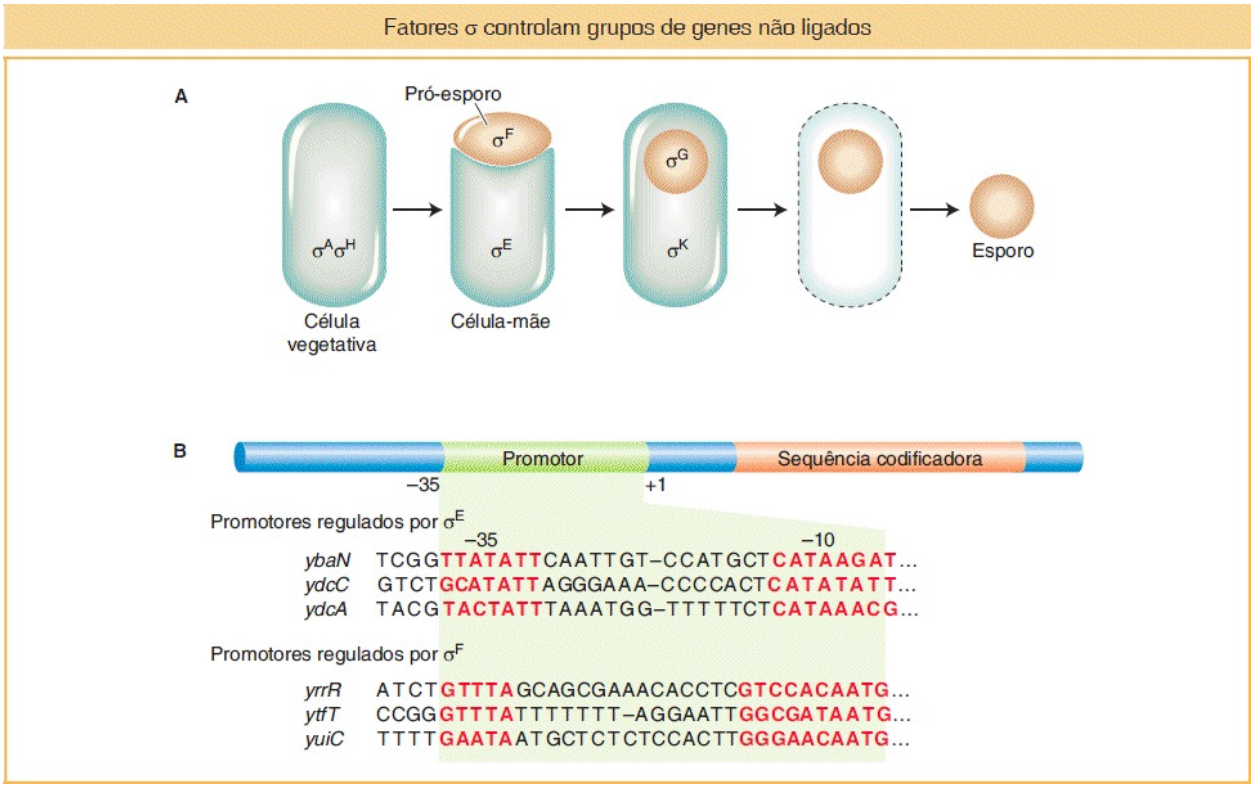


FIGURA 11.31 A esporulação no *Bacillus subtilis* é regulada por cascatas de fatores σ. **A.** Em células vegetativas, σ^A e σ^H são ativos. No início da esporulação, σ^F é ativo no pró-esporo e σ^E é ativo na célula-mãe. Esses fatores σ em seguida são substituídos por σ^G e σ^K, respectivamente. A célula-mãe finalmente é lisada e libera o esporo maduro. **B.** Os fatores σ^E e σ^F controlam os regulons de muitos genes (*ybaN*, e assim por diante, nesta ilustração). Estão demonstrados três exemplos do grande número de promotores regulados por cada fator σ. Cada fator σ apresenta uma sequência de ligação distinta e específica nas sequências —35 e —10 dos promotores-alvo. (Dados de P. Eichenberger *et al.*, *J. Mol. Biol.* 327, 2003, 945-972; e S. Wang *et al.*, *J. Mol. Biol.* 358, 2006, 16-37.)

Como esses fatores σ alternativos controlam diferentes aspectos do processo de esporulação? As respostas se tornaram cristalinas com o advento de novas abordagens para a caracterização da expressão de todos os genes em um genoma (ver Seção 14.6). Atualmente, é possível monitorar a transcrição de cada gene de *B. subtilis* durante o crescimento vegetativo e a formação de esporos e em

diferentes compartimentos do esporo. Diversas centenas de genes foram identificadas desse modo, os quais são transcricionalmente ativados ou reprimidos durante a formação dos esporos.

Como os diferentes conjuntos de genes são controlados por cada fator σ? Cada fator σ apresenta diferentes propriedades de ligação sequência específica ao DNA. Os óperons ou genes individuais regulados por fatores σ específicos apresentam sequências caracterizadas nas regiões —35 e —10 de seus promotores que são ligadas por um fator σ e não por outros (Figura 11.31 B). Por exemplo, σ^E se liga a no mínimo 121 promotores, dentro de 34 óperons e 87 genes individuais, para regular mais de 250 genes, e σ^F se liga a no mínimo 36 promotores para regular 48 genes.

> CONCEITO-CHAVE A expressão sequencial de fatores σ alternativos que reconhecem sequências promotoras alternativas proporciona a expressão coordenada de grandes números de óperons independentes e genes não ligados durante o programa de desenvolvimento da esporulação.

Fatores σ alternativos também desempenham papéis importantes na virulência de patógenos humanos. Por exemplo, bactérias do gênero *Clostridium* produzem toxinas potentes que são responsáveis por diversas doenças, tais como botulismo, tétano e gangrena. Recentemente descobriu-se que os genes de toxinas-chave de *C. botulinum, C. tetani* e *C. perfringens* são controlados por fatores σ alternativos correlatos, que reconhecem sequências semelhantes nas regiões —35 e —10 dos genes de toxina. A compreensão sobre os mecanismos de regulação dos genes de toxina pode levar a novos meios de prevenção de doenças e terapia.

RESUMO

A regulação gênica com frequência é mediada por proteínas que reagem a sinais ambientais por meio da elevação ou da redução das taxas de transcrição de genes específicos. A lógica dessa regulação é direta. Para que a regulação opere de

modo adequado, as proteínas reguladoras apresentam sensores integrados que monitoram continuamente as condições celulares. As atividades dessas proteínas dependem então de condições ambientais apropriadas.

Em bactérias e seus vírus, o controle de diversos genes estruturais pode ser coordenado pelo agrupamento dos genes em óperons no cromossomo, de modo que eles são transcritos em mRNA multigênicos. O controle coordenado simplifica a tarefa para as bactérias, tendo em vista que um grupo de sítios reguladores por óperon é suficiente para regular a expressão de todos os genes do óperon. Alternativamente, o controle coordenado também pode ser conquistado por meio de fatores σ discretos que regulam dúzias de promotores independentes simultaneamente.

No controle regulador negativo, uma proteína repressora bloqueia a transcrição por meio da ligação ao DNA no sítio operador. O controle regulador negativo é exemplificado pelo sistema *lac*. A regulação negativa é um modo muito direto de o sistema *lac* desligar genes na ausência de açúcares apropriados no ambiente. No controle regulador positivo, são necessários fatores proteicos para ativar a transcrição. Alguns controles gênicos procarióticos, tais como aquele em relação à repressão catabólica, operam por meio do controle gênico positivo.

Muitas proteínas reguladoras são membros de famílias de proteínas que apresentam *motifs* de ligação ao DNA muito semelhantes, tais como o domínio hélice-giro-hélice. Outras partes das proteínas, tais como seus domínios de interação proteína-proteína, tendem a ser menos semelhantes. A especificidade da regulação gênica depende de interações químicas das cadeias laterais dos aminoácidos com grupos químicos nas bases do DNA.

TERMOS-CHAVE

ação cis
adenosina monofosfato cíclico (cAMP)
antiterminalizador
atenuação
atenuador

ativador
ciclo lisogênico
ciclo lítico
controle negativo
controle positivo
diploide parcial
domínio de ligação ao DNA
efetor alostérico
genes controlados coordenadamente
indução
indutor
iniciador
interruptor genético
mutação constitutiva
operador
óperon
promotor
proteína ativadora de catabólito (CAP)
regulon
repressão catabólica
repressor
sequência líder
sítio alostérico
transatuante
transição alostérica

PROBLEMAS RESOLVIDOS

Este conjunto de quatro problemas resolvidos, que são semelhantes ao Problema 10 nos Problemas básicos, ao fim deste capítulo, é desenhado para testar a compreensão sobre o modelo do óperon. Aqui, recebemos diversos diploides e nos é solicitado que determinemos se os produtos dos genes Z e Y são produzidos

na presença ou na ausência de um indutor. Utilize uma tabela semelhante àquela no Problema 11 como base para as suas respostas, com a exceção de que os títulos das colunas serão como segue:

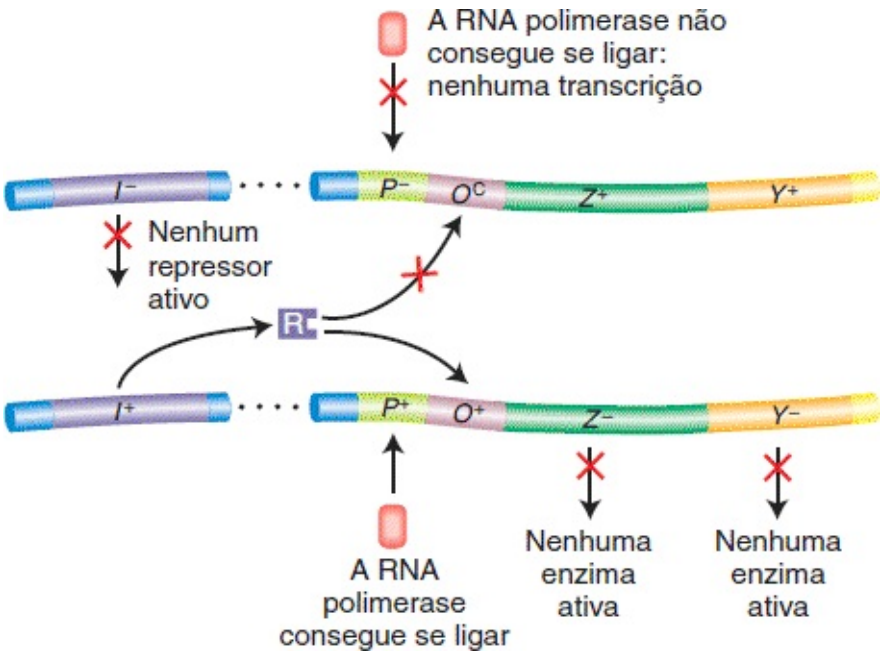
	Gene Z		*Gene Y*	
Genótipo	Nenhum indutor	Indutor	Nenhum indutor	Indutor

PROBLEMA RESOLVIDO 1

$$\frac{I^- P^- O^C Z^+ Y^+}{I^+ P^+ O^+ Z^- Y^-}$$

Solução

Um modo de abordar estes problemas é primeiramente considerar cada cromossomo em separado e em seguida construir um diagrama. A ilustração a seguir digrama este diploide:



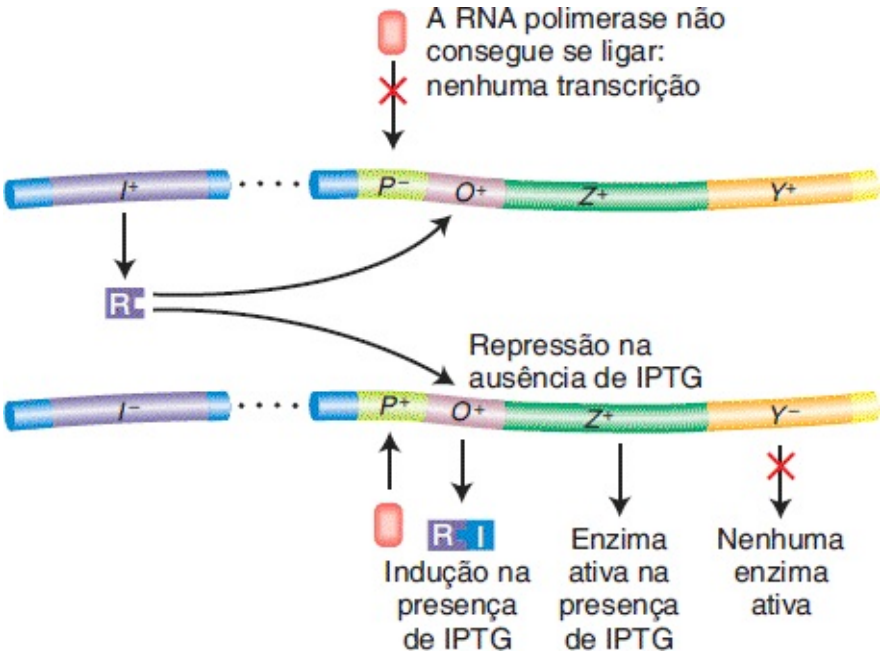
O primeiro cromossomo é P^- e, assim, a transcrição é bloqueada e nenhuma enzima Lac pode ser sintetizada a partir dele. O segundo cromossomo (P^+) pode ser transcrito e, portanto, a transcrição é reprimível (O^+). Entretanto, os genes estruturais ligados ao bom promotor são defeituosos; portanto, nenhum produto de *Z* ou produto de *Y* ativo pode ser produzido. Os símbolos a serem adicionados à sua tabela são "—, —, —, —".

PROBLEMA RESOLVIDO 2

$$\frac{I^+\ P^-\ O^+\ Z^+\ Y^+}{I^-\ P^+\ O^+\ Z^+\ Y^-}$$

Solução

O primeiro cromossomo é P^- e, assim, nenhuma enzima pode ser sintetizada a partir dele. O segundo cromossomo é O^+ e, assim, a transcrição é reprimida pelo repressor fornecido a partir do primeiro cromossomo, que pode atuar em trans através do citoplasma. Entretanto, apenas o gene *Z* desse cromossomo está intacto. Portanto, na ausência de um indutor, nenhuma enzima é produzida; na presença de um indutor, apenas o produto do gene *Z*, a β-galactosidase, é produzida. Os símbolos a serem adicionados à tabela são "—, +, —, —,".

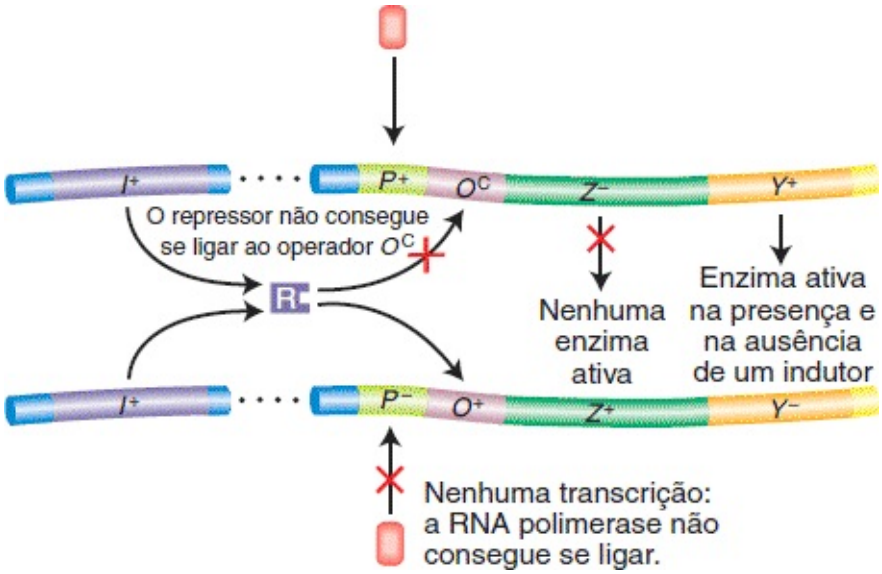


PROBLEMA RESOLVIDO 3

$$\frac{I^+\ P^+\ O^C\ Z^-\ Y^+}{I^+\ P^-\ O^+\ Z^+\ Y^-}$$

Solução

Tendo em vista que o segundo cromossomo é P^-, precisamos considerar apenas o primeiro cromossomo. Esse cromossomo é O^C e, assim, a enzima é produzida na ausência de um indutor, muito embora, em virtude da mutação Z^-, apenas a permease ativa (Y) seja produzida. As entradas na tabela devem ser "—, —, +, +".

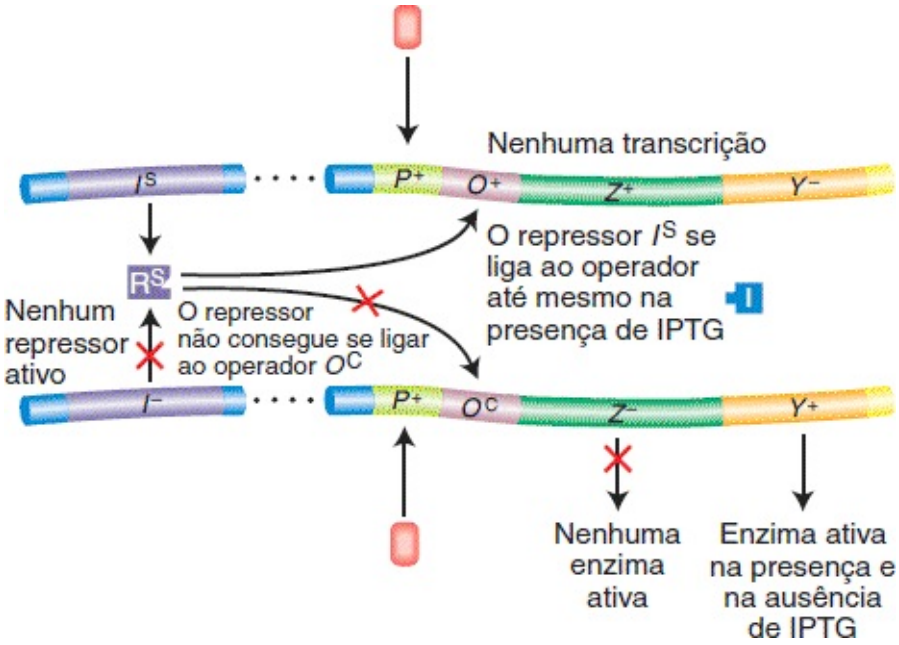


PROBLEMA RESOLVIDO 4

$$\frac{I^S\ P^+\ O^+\ Z^+\ Y^-}{I^-\ P^+\ O^C\ Z^-\ Y^+}$$

Solução

Na ausência de um repressor I^S, todos os operadores do tipo selvagem são desligados, tanto com um indutor quanto sem um indutor. Portanto, o primeiro cromossomo é incapaz de produzir qualquer enzima. Entretanto, o segundo cromossomo apresenta um operador alterado (O^C) e pode produzir enzima tanto na ausência quanto na presença de um indutor. Apenas o gene *Y* é do tipo selvagem no cromossomo O^C e, assim, apenas a permease é produzida constitutivamente. As entradas na tabela devem ser "—, —, +, +".



PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Compare a estrutura de IPTG demonstrada na Figura 11.7 com a estrutura da galactose demonstrada na Figura 11.5. Por que o IPTG se liga ao repressor Lac, mas não é fragmentado pela β-galactosidase?

2. Observando a Figura 11.9, por que os diploides parciais eram essenciais para estabelecer a natureza de ação trans do repressor Lac? Seria possível distinguir os genes de ação cis dos genes de ação trans (transatuantes) em haploides?

3. Por que as mutações no promotor se agrupam nas posições —10 e —35, conforme demonstrado na Figura 11.11? Qual interação de proteína-DNA é rompida por essas mutações?

4. Observando a Figura 11.16, verifique a grande sobreposição entre o operador e a região do óperon *lac* que é transcrita. Qual proteína se liga especificamente a essa sequência sobreposta, e qual efeito ela apresenta sobre a transcrição?

5. Examinando a Figura 11.21, qual efeito você prevê que as mutações do

trpA terão sobre os níveis de triptofano?

6. Examinando a Figura 11.21, qual efeito você prevê que as mutações do *trpA* terão sobre a expressão do mRNA do *trp*?

PROBLEMAS BÁSICOS

7. Qual das moléculas a seguir é um indutor do óperon lac:

a. Galactose

b. Glicose

c. Alolactose

d. Isotiocianato

e. cAMP

f. Lactose

8. Explique por que os alelos I^- no sistema *lac* normalmente são recessivos aos alelos I^+ e por que os alelos I^+ são recessivos aos alelos I^S.

9. O que queremos dizer quando falamos que as mutações O^C no sistema *lac* são de ação cis?

10. Os símbolos *a*, *b* e *c* na tabela a seguir representam os genes do sistema *lac* de *E. coli* em relação ao repressor (*I*), à região operadora (*O*) e à β-galactosidase (*Z*), embora não necessariamente nessa ordem. Além disso, a ordem na qual os símbolos estão escritos nos genótipos não é necessariamente a sequência real no óperon *lac*.

Atividade (+) ou inatividade (—) do gene *Z*

Genótipo	Indutor ausente	Indutor presente
$a^- b^+ c^+$	+	+
$a^+ b^+ c^-$	+	+

$a^+ b^- c^-$	—	—
$a^+ b^- c^+/a^- b^+ c^-$	+	+
$a^+ b^+ c^+/a^- b^- c^-$	—	+
$a^+ b^+ c^-/a^- b^- c^+$	—	+
$a^- b^+ c^+/a^+ b^- c^-$	+	+

a. Qual símbolo (*a, b* ou *c*) representa cada um dos genes *lac I, O* e *Z*?
b. Na tabela, um sinal de menos sobrescrito em um símbolo de gene indica meramente um mutante, mas alguns comportamentos mutantes nesse sistema recebem designações de mutantes especiais. Com a utilização dos símbolos gênicos convencionais em relação ao óperon lac, designe cada genótipo na tabela.

11. O mapa do óperon *lac* é:

POZY

A região promotora (*P*) é o local de início da transcrição por meio da ligação da molécula de RNA polimerase antes da real produção de mRNA. Promotores alterados por mutação (P^-) aparentemente não conseguem ligar a molécula de RNA polimerase. Podem ser feitas determinadas previsões a respeito do efeito das mutações P^-. Utilize as suas previsões e o seu conhecimento sobre o sistema lactose para completar a tabela a seguir. Insira um "+" onde uma enzima é produzida e um "—" onde nenhuma enzima é produzida. A primeira foi realizada como um exemplo.

12. Explique as diferenças fundamentais entre o controle negativo e o controle positivo da transcrição em procariotos. Cite dois exemplos de cada

mecanismo de controle.

13. Mutantes que são *lacY*$^{-}$ retêm a capacidade de sintetizar β-galactosidase. Entretanto, embora o gene *lacI* ainda esteja intacto, a β-galactosidade deixou de ser induzida pela adição de lactose ao meio. Explique.

	β-galactosidase		*Permease*	
Genótipo	Nenhuma lactose	Lactose	Nenhuma lactose	Lactose
$\frac{I^+ P^+ O^+ Z^+ Y^+}{I^+ P^+ O^+ Z^+ Y^+}$	–	+	–	+
a. $\frac{I^- P^+ O^C Z^+ Y^-}{I^+ P^+ O^+ Z^- Y^+}$				
b. $\frac{I^+ P^- O^C Z^- Y^+}{I^- P^+ O^C Z^+ Y^-}$				
c. $\frac{I^S P^+ O^+ Z^+ Y^-}{I^+ P^+ O^+ Z^- Y^+}$				
d. $\frac{I^S P^+ O^+ Z^+ Y^+}{I^- P^+ O^+ Z^+ Y^+}$				
e. $\frac{I^- P^+ O^C Z^+ Y^-}{I^- P^+ O^+ Z^- Y^+}$				
f. $\frac{I^- P^- O^+ Z^+ Y^+}{I^- P^+ O^C Z^+ Y^-}$				
g. $\frac{I^+ P^+ O^+ Z^- Y^+}{I^- P^+ O^+ Z^+ Y^-}$				

14. Quais são as analogias entre os mecanismos que controlam o óperon *lac* e aqueles que controlam os interruptores genéticos do fago λ?

15. Compare o arranjo dos sítios de ação cis nas regiões de controle do óperon *lac* e do fago λ.

16. Qual proteína reguladora induz os genes da fase lítica do ciclo de vida do fago λ?

a. cI

b. Cro

c. Repressor Lac

d. Lactose

17. Preveja o efeito de uma mutação que elimine a atividade de ligação ao DNA da proteína σ^E na formação de esporos no *Bacillus subtilis*.

PROBLEMAS DESAFIADORES

18. Uma mutação interessante em *lacI* resulta em repressores com aumento de 110 vezes de ligação a ambos o DNA operador e não operador. Esses repressores demonstram uma curva de indução "reversa", que possibilita a síntese de β-galactosidase na ausência de um indutor (IPTG), mas que reprime parcialmente a expressão de β-galactosidase na presença de IPTG. Como você pode explicar isso? (Observe que, quando o IPTG se liga a um repressor, ele não destrói completamente a afinidade pelo operador, mas em vez disso, reduz a afinidade em 110 vezes. Adicionalmente, na medida em que as células se dividem e novos operadores são produzidos pela síntese de filamentos-filhos, o repressor deve encontrar os novos operadores por meio da busca ao longo do DNA, ligando-se rapidamente a sequências não operadoras e se dissociando delas.)

19. Determinadas mutações *lacI* eliminam a ligação do operador ao repressor Lac, mas não afetam a agregação de subunidades para a produção de um tetrâmero, a forma ativa do repressor. Essas mutações são parcialmente dominantes em relação ao tipo selvagem. Você pode explicar o fenótipo I^- parcialmente dominante dos heterodiploides I^-/I^+?

20. Você está examinando a regulação do óperon da lactose na bactéria *Escherichia coli*. Você isola sete novas linhagens mutantes independentes que não apresentam os produtos de todos os três genes estruturais. Você suspeita que algumas dessas mutações são $lacI^S$ e que outras mutações são alterações que evitam a ligação da RNA polimerase à região promotora. Com a utilização de quaisquer genótipos haploides e diploides parciais que você considere necessários, descreva um conjunto de genótipos que permitirá que você diferencie entre as classes *lacI* e *lacP* de mutações não

indutíveis.

21. Você está estudando as propriedades de um novo tipo de mutação reguladora do óperon da lactose. Essa mutação, denominada *S*, leva à repressão completa dos genes *lacZ, lacY* e *lacA*, independentemente da presença da lactose. Os resultados dos estudos dessa mutação em diploides parciais demonstram que essa mutação é completamente dominante sobre o tipo selvagem. Quando você trata bactérias da linhagem mutante *S* com um mutágeno e seleciona bactérias mutantes que conseguem expressar as enzimas codificadas pelos genes *lacZ, lacY* e *lacA* na presença da lactose, algumas mutações mapeiam na região do operador *lac* e outras para o gene repressor do *lac*. Com base no seu conhecimento sobre o óperon da lactose, forneça uma explicação genética molecular para todas essas propriedades da mutação *S*. Inclua uma explicação sobre a natureza constitutiva das "mutações reversas".

22. O óperon *trp* na *E. coli* codifica enzimas essenciais para a biossíntese do triptofano. O mecanismo geral para o controle do óperon *trp* é semelhante àquele observado com o óperon *lac*: quando o repressor se liga ao operador, a transcrição é evitada; quando o repressor não se liga (ao operador), a transcrição prossegue. A regulação do óperon *trp* difere da regulação do óperon *lac* da seguinte maneira: as enzimas codificadas pelo óperon *trp* não são sintetizadas quando o triptofano está presente, mas sim ausente. No óperon *trp*, o repressor apresenta dois sítios de ligação: um para o DNA e o outro para a molécula efetora, o triptofano. O repressor *trp* deve primeiramente se ligar a uma molécula de triptofano antes que possa efetivamente se ligar ao operador *trp*.

a. Desenhe um mapa do óperon do triptofano, indicando o promotor (*P*), o operador (*O*) e o primeiro gene estrutural do óperon do triptofano (*trpA*). No seu desenho, indique onde no DNA a proteína repressora se liga quando está ligada ao triptofano.

b. O gene *trpR* codifica o repressor; *trpO* é o operador; *trpA* codifica a enzima triptofano sintetase. Um repressor $trpR^{-}$ não consegue se ligar ao

triptofano, um operador *trpO*$^{-}$ não consegue se ligar ao repressor, e a enzima codificada por um gene mutante *trpA*$^{-}$ é completamente inativa. Você espera encontrar uma triptofano sintetase ativa em cada uma das linhagens mutantes a seguir quando as células são cultivadas na presença de triptofano? Na sua ausência?

(1) $R^{+}\ O^{+}\ A^{+}$ (tipo selvagem)

(2) $R^{-}\ O^{+}\ A^{+}/R^{+}\ O^{+}\ A^{-}$

(3) $R^{+}\ O^{-}\ A^{+}/R^{+}\ O^{+}\ A^{-}$

23. É medida a atividade da enzima β-galactosidase produzida por células do tipo selvagem cultivadas em meio suplementado com diferentes fontes de carbono. Em unidades relativas, são observados os níveis de atividade a seguir:

Glicose	Lactose	Lactose + glicose
0	100	1

Preveja os níveis relativos de atividade da β-galactosidase em células cultivadas sob condições semelhantes quando as células são *lacI*$^{-}$, *lacI*^{S}, *lacO*$^{+}$ e *crp*$^{-}$.

24. Observa-se que um bacteriófago λ é capaz de lisogenizar sua *E. coli* hospedeira a 30°C, mas não a 42°C. Quais genes podem ser mutantes nesse fago?

25. O que aconteceria com a capacidade do bacteriófago λ de lisar uma célula hospedeira se ela adquirisse uma mutação no sítio de ligação O_R em relação à proteína Cro? Por quê?

26. Contraste os efeitos das mutações nos genes que codificam os fatores σ de esporulação específica com as mutações nas regiões —35 e —10 dos promotores de genes em seus regulons.

a. As mutações funcionais nos genes do fator σ ou nos promotores

individuais apresentariam o maior efeito sobre a esporulação?

b. Com base nas sequências demonstradas na Figura 11.30 B, você esperaria que todas as mutações de ponto nas regiões —35 ou —10 afetassem a expressão gênica?

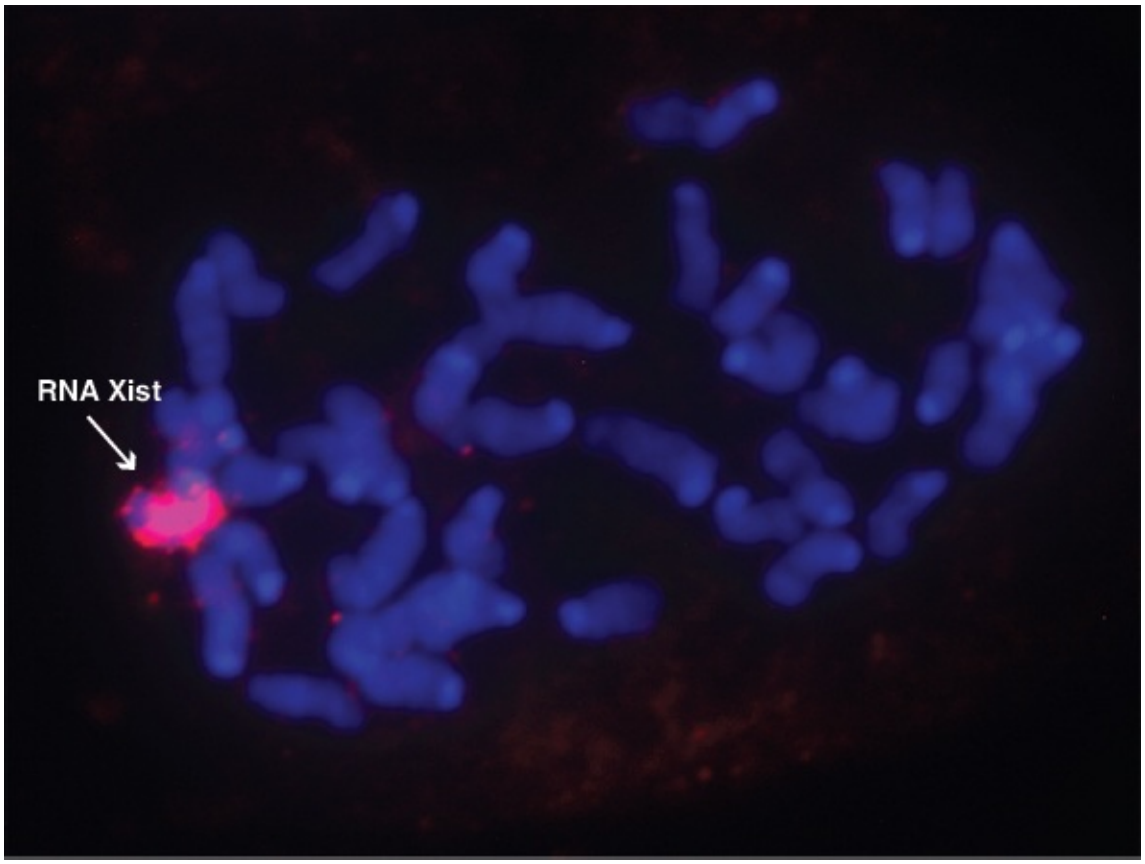
[1]F. Jacob, *The Statue Within: An Autobiography*, 1988.

[2]H. F. Judson, *The Eight Day of Creation: Makers of the Revolution in Biology*, 1979.

[3]F. Jacob e J. Monod, *Cold Spring Harbor Quant. Symp. Biol.* 26, 1963, 393.

CAPÍTULO 12

Regulação da Expressão Gênica em Eucariotos



O RNA Xist (marcado por um corante rodamina vermelho) abrange uma das duas cópias do cromossomo X. A expressão de Xist levará à inativação do cromossomo. A imagem é de um experimento de hibridização fluorescente *in situ* (FISH) de RNA realizado em um cromossomo em metáfase obtido a partir de uma linhagem celular de fibroblasto feminino. (J. T. Lee *et al.*, "Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome", *Genes Dev.*, 23 (16), 2009, 1831-1842, Fig. 2. © Cold Spring Harbor Laboratory Press. Fotografia de Jeannie Lee.)

TÓPICOS

12.1 Regulação da transcrição em eucariotos | Visão geral

12.2 Lições das leveduras | Sistema GAL
12.3 Cromatina dinâmica
12.4 Ativação de genes em um ambiente de cromatina
12.5 Inativação a longo prazo de genes em um ambiente de cromatina
12.6 Silenciamento gênero-específico de genes e cromossomos inteiros
12.7 Repressão gênica pós-transcricional pelos miRNA

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Comparar e contrastar os mecanismos moleculares da regulação gênica em eucariotos e em bactérias
- Explicar como os eucariotos geram muitos padrões diferentes de expressão gênica com um número limitado de proteínas reguladoras
- Discutir o envolvimento da cromatina na regulação gênica eucariótica
- Descrever o conceito de marcas epigenéticas e discutir como podem atuar no DNA e nas proteínas
- Comparar e contrastar os papéis desempenhados por moléculas de RNA na repressão da expressão gênica eucariótica.

A clonagem de Dolly, uma ovelha, foi relatada mundialmente em 1996 (Figura 12.1). Dolly se desenvolveu a partir de núcleos somáticos adultos que haviam sido implantados dentro de ovócitos enucleados (ovócitos com os núcleos removidos). Mais recentemente, vacas, porcos, camundongos e outros mamíferos também foram clonados com a utilização de tecnologias semelhantes. A clonagem bem-sucedida de Dolly foi uma grande surpresa para a comunidade científica, tendo em vista que sabidamente a formação de gametas (espermatozoides e ovócitos) inclui modificações específicas do sexo nos respectivos genomas, que resultam em padrões de expressão gênica específicos do sexo. Dolly é um símbolo do quanto progredimos na compreensão dos aspectos da regulação gênica eucariótica, tais como o controle global da expressão gênica exemplificado pelo desenvolvimento dos gametas. Entretanto, para cada clone de sucesso, incluindo Dolly, existem muitos mais, talvez centenas

de embriões, que falham em se desenvolver em progênie viável. A taxa de insucesso extremamente alta enfatiza o quanto ainda falta para a ser decifrado a respeito da regulação gênica eucariótica.

Neste capítulo, examinaremos a regulação gênica em eucariotos. De muitos modos, a nossa observação da regulação gênica será um estudo de contrastes. No Capítulo 11, você aprendeu como as atividades dos interruptores genéticos em bactérias com frequência eram reguladas por proteínas ativadoras ou repressoras únicas e como o controle de grupos de genes foi conquistado por meio de sua organização em óperons ou por meio da atividade de fatores específicos. As expectativas iniciais eram de que a expressão gênica eucariótica seria regulada por meios semelhantes. Em eucariotos, entretanto, a maior parte dos genes não é encontrada em óperons. Além disso, veremos que as proteínas e as sequências de DNA que participam na regulação gênica eucariótica são mais numerosas. Com frequência, muitas proteínas de ligação ao DNA atuam sobre um interruptor único, com muitos interruptores separados por gene, e as sequências reguladoras desses interruptores com frequência estão localizadas longe dos promotores. Mais uma diferença-chave entre as bactérias e os eucariotos é que o acesso aos promotores gênicos eucarióticos é restringido pela cromatina. A regulação gênica em eucariotos requer a atividade de grandes complexos proteicos que promovem ou restringem o acesso a promotores gênicos pela RNA polimerase. Este capítulo proporcionará um fundamento essencial para a compreensão da regulação espaço-temporal da expressão gênica que realiza a coreografia do processo de desenvolvimento descrito no Capítulo 13.

O primeiro mamífero clonado

FIGURA 12.1 O primeiro mamífero clonado foi uma ovelha chamada Dolly. (© Roslin Institute/Phototake.)

12.1 Regulação da transcrição em eucariotos | Visão geral

As propriedades biológicas de cada tipo de célula eucariótica são amplamente determinadas pelas proteínas expressas dentro dela. Essa constelação de proteínas expressas determina uma grande parte da arquitetura da célula, suas atividades enzimáticas, suas interações com o seu ambiente e muitas outras propriedades fisiológicas. Entretanto, em qualquer momento determinado na história de vida de uma célula, apenas uma fração dos RNA e das proteínas codificadas em seu genoma é expressa. Em momentos diferentes, o perfil dos produtos de genes expressos pode diferir dramaticamente, tanto em relação a quais proteínas são expressas quanto em quais níveis. Como esses perfis específicos são gerados?

Assim como poderia ser esperado, se o produto final é uma proteína, a regulação pode ser alcançada por meio do controle da transcrição do DNA em

RNA ou da tradução do RNA em proteína. De fato, a regulação gênica ocorre em muitos níveis, incluindo o nível do mRNA (por meio de alterações na recomposição ou na estabilidade do mRNA) e após a tradução (por meio de modificações das proteínas). Os diversos modos como os genes eucarióticos podem ser regulados foram divididos em duas categorias gerais que refletem quando eles atuam: regulação gênica transcricional e regulação gênica pós-transcricional. Enquanto a primeira será o enfoque primário deste capítulo, a última está recebendo cada vez mais atenção. Em particular, o papel do RNA de atuar na repressão pós-transcricional da expressão gênica (denominado silenciamento gênico; ver Capítulo 8) é uma das áreas de maior interesse das pesquisas atuais. Três dos tipos de RNA que participam no silenciamento gênico, o miRNA, o ncRNA e o siRNA, foram introduzidos no Capítulo 8. Posteriormente neste capítulo, exploraremos os mecanismos de regulação gênica mediada pelo miRNA e pelo ncRNA.

A maior parte da regulação caracterizada até o momento ocorre no nível da transcrição gênica; portanto, neste capítulo, o foco primário é na regulação da transcrição gênica. O mecanismo básico em atuação é que sinais moleculares externos ou internos à célula levam à ligação de proteínas reguladoras a sítios específicos do DNA fora das regiões codificadoras de proteínas e a ligação dessas proteínas modula a velocidade da transcrição. Essas proteínas podem auxiliar direta ou indiretamente a RNA polimerase na ligação ao seu sítio de início da transcrição – o *promotor* – ou podem reprimir a transcrição ao evitar a ligação da RNA polimerase.

Embora bactérias e eucariotos compartilhem uma grande parte da lógica da regulação gênica, existem algumas diferenças fundamentais nos mecanismos subjacentes e no maquinário. Ambos utilizam proteínas de ligação ao DNA com sequências específicas para modular o nível de transcrição. Entretanto, os genomas eucarióticos são maiores e sua diversidade de propriedades é maior do que aquela das bactérias. Inevitavelmente, a regulação dos genomas eucarióticos é mais complexa, refletindo a sua complexidade estrutural e funcional. Isso requer mais tipos de proteínas reguladoras e mais tipos de interações com as regiões reguladoras adjacentes no DNA. Uma diferença notável é que o DNA eucariótico

é compactado em *nucleossomos*, formando a *cromatina*, enquanto o DNA bacteriano não apresenta nucleossomos. Em eucariotos, a estrutura da cromatina é dinâmica e é um ingrediente essencial na regulação gênica.

Em geral, o estado básico de um gene bacteriano é "ligado". Portanto, normalmente a RNA polimerase consegue se ligar diretamente a um promotor quando nenhuma outra proteína reguladora está próxima para se ligar ao DNA. Em bactérias, o início da transcrição é evitado ou reduzido se a ligação da RNA polimerase estiver bloqueada, normalmente por meio da ligação de uma proteína reguladora repressora. As proteínas reguladoras ativadoras aumentam a ligação da RNA polimerase aos promotores onde pouco auxílio é necessário. Contrariamente, o estado básico em eucariotos é "desligado". Portanto, o maquinário de transcrição (incluindo a RNA polimerase II e os fatores gerais de transcrição associados) não consegue se ligar ao promotor na ausência de outras proteínas reguladoras (Figura 12.2). Em muitos casos, não é possível a ligação do aparato transcricional, tendo em vista que nucleossomos estão posicionados para bloquear o promotor. Portanto, a estrutura da cromatina normalmente precisa ser alterada para ativar a transcrição eucariótica. Aquelas alterações em geral dependem da ligação de proteínas reguladoras a uma sequência específica do DNA. A estrutura da cromatina ao redor dos genes ativados ou reprimidos dentro das células pode ser consideravelmente estável e herdada pelas células-filhas. A herança dos estados da cromatina é um tipo de herança que não envolve diretamente a sequência de DNA; ela proporciona um meio de regulação *epigenética*.

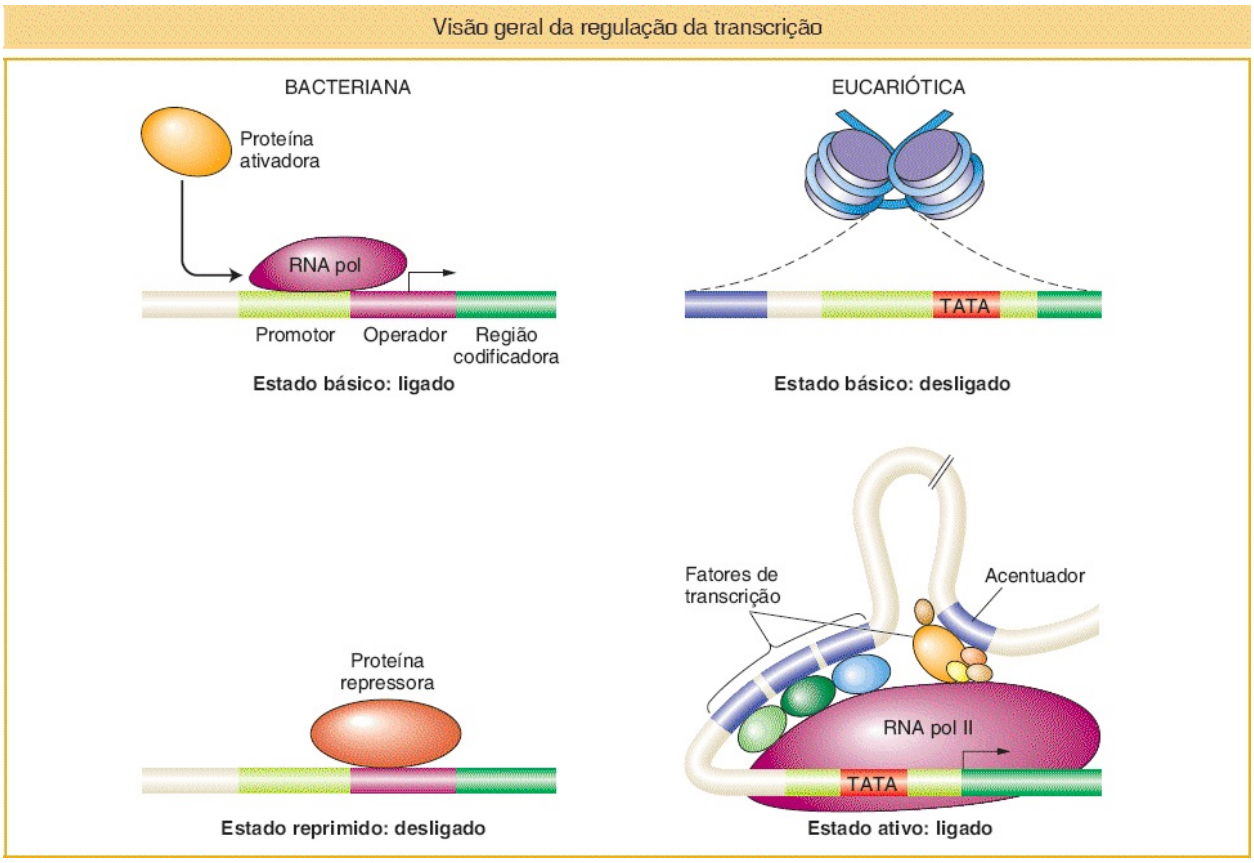


FIGURA 12.2 Em bactérias, a RNA polimerase normalmente consegue iniciar a transcrição, exceto se uma proteína repressora a bloquear. Em eucariotos, entretanto, a compactação do DNA com os nucleossomos evita a transcrição, exceto se outras proteínas reguladoras estiverem presentes. Essas proteínas reguladoras expõem as sequências promotoras ao alterar a densidade ou a posição do nucleossomo. Elas também podem recrutar a RNA polimerase II mais diretamente por meio de ligação.

As características únicas da regulação transcricional eucariótica são o foco deste capítulo. Algumas diferenças das bactérias no processo da transcrição e que afetam a regulação já foram observadas no Capítulo 8:

1. Em bactérias, todos os genes são transcritos em RNA pela mesma RNA polimerase, enquanto três RNA polimerases atuam em eucariotos. A RNA polimerase II, que transcreve DNA em mRNA, foi o foco do Capítulo 8 e será a única polimerase discutida neste capítulo.
2. Os transcritos de RNA são extensivamente processados durante a transcrição em eucariotos; as extremidades 5′ e 3′ são modificadas e os íntrons são removidos.

3. A RNA polimerase II é muito maior e mais complexa do que a sua correspondente bacteriana. Um motivo de ser mais complexa é que a RNA polimerase II deve sintetizar RNA *e* coordenar os eventos únicos de processamento especiais de eucariotos.

Os eucariotos multicelulares podem apresentar até 25.000 genes, diversas vezes mais do que a bactéria média. Além disso, os padrões de expressão gênica eucariótica podem ser extraordinariamente complexos. Ou seja, o momento da expressão gênica e a quantidade de transcritos produzida variam amplamente entre os genes eucarióticos. Por exemplo, um gene pode ser transcrito apenas durante um estágio do desenvolvimento e outro apenas na presença de uma infecção viral. Finalmente, a maioria dos genes em uma célula eucariótica está desligada em qualquer momento. Com base nessas considerações isoladamente, a regulação gênica eucariótica deve ser capaz de

1. Assegurar que a expressão da maior parte dos genes no genoma esteja desligada em qualquer momento enquanto ocorre a ativação de um subgrupo de genes.
2. Gerar milhares de padrões de expressão gênica.

Conforme você verá posteriormente no capítulo, foram desenvolvidos mecanismos para assegurar que a maior parte dos genes em uma célula eucariótica não seja transcrita. Antes de considerar como os genes são mantidos transcricionalmente inativos, enfocaremos o segundo ponto: como os genes eucarióticos são capazes de exibir enorme número e diversidade de padrões de expressão? O maquinário necessário para a geração de tantos padrões de transcrição gênica *in vivo* apresenta muitos componentes, incluindo as *proteínas* reguladoras de ação trans e as *sequências de DNA* reguladoras de ação cis. Podemos dividir as proteínas reguladoras em dois conjuntos com base nas sequências reguladoras do DNA a que elas se ligam. O primeiro conjunto de proteínas é o grande complexo de RNA polimerase II e os fatores gerais de transcrição a respeito dos quais você aprendeu no Capítulo 8. Para iniciar a transcrição, essas proteínas interagem com sequências reguladoras de ação cis no

DNA denominadas **elementos proximais do promotor**, próximas do promotor de um gene. O segundo conjunto de proteínas reguladoras é composto por fatores de transcrição que se ligam às sequências reguladoras de ação cis no DNA denominadas **acentuadores**. Essas sequências reguladoras podem estar localizadas a uma distância considerável dos promotores gênicos. Falando em termos gerais, os promotores e seus elementos proximais são ligados por fatores de transcrição que afetam a expressão de muitos genes. Os acentuadores são ligados por fatores de transcrição que controlam a regulação de subconjuntos menores de genes. Com frequência, um acentuador atuará em apenas um ou em alguns poucos tipos celulares em um eucarioto multicelular. Uma grande parte da estratégia do controle transcricional eucariótico depende de como fatores de transcrição específicos controlam o acesso dos fatores gerais de transcrição e da RNA polimerase II.

Para que a RNA polimerase II transcreva o DNA em RNA a uma velocidade máxima, múltiplos elementos reguladores de ação cis devem participar. Os promotores, seus elementos proximais e os acentuadores são, todos, alvos para a ligação ao DNA de diferentes proteínas de ação trans. A Figura 12.3 é uma representação esquemática do promotor e dos elementos da sequência proximal do promotor. A ligação da RNA polimerase II ao promotor não produz transcrição eficiente por si própria. A transcrição requer a ligação de fatores gerais de transcrição a elementos proximais do promotor, que são comumente encontrados dentro de 100 pb do sítio de início da transcrição de muitos genes (mas não de todos). Um desses elementos é o CCAAT boxe (pronunciado “cat”) e outro geralmente é um segmento rico em GC mais distante *upstream*. Os fatores gerais de transcrição que se ligam aos elementos proximais do promotor são expressos na maior parte das células e, assim, estão disponíveis para iniciar a transcrição a qualquer momento. Mutações nesses sítios podem apresentar um efeito dramático na transcrição, demonstrando quão importantes eles são. Se esses elementos de sequência estiverem mutados, o nível de transcrição em geral é reduzido, conforme demonstrado na Figura 12.4.

Para modular a transcrição, as proteínas reguladoras apresentam um ou mais dos *domínios funcionais* a seguir:

1. Um domínio que reconhece uma sequência reguladora no DNA (o sítio de ligação da proteína ao DNA).
2. Um domínio que interage com uma ou mais proteínas do aparato de transcrição (RNA polimerase ou uma proteína associada à RNA polimerase).

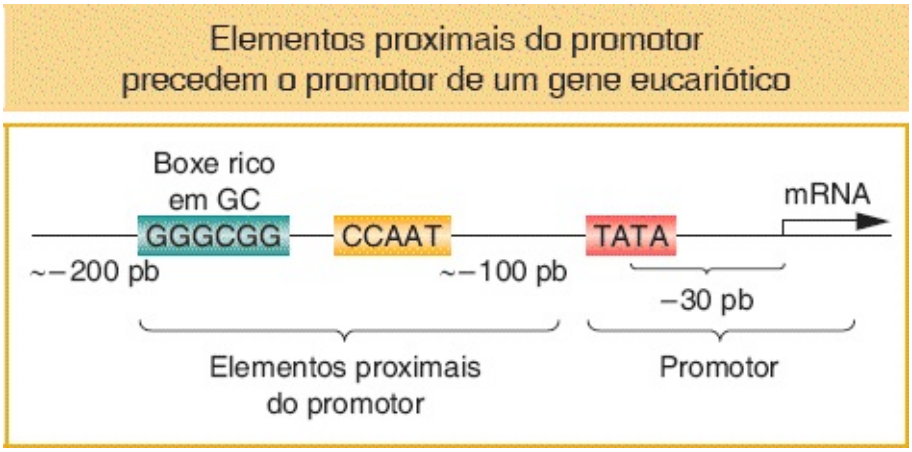


FIGURA 12.3 A região *upstream* do sítio de início da transcrição em eucariotos superiores contém elementos proximais do promotor e o promotor.

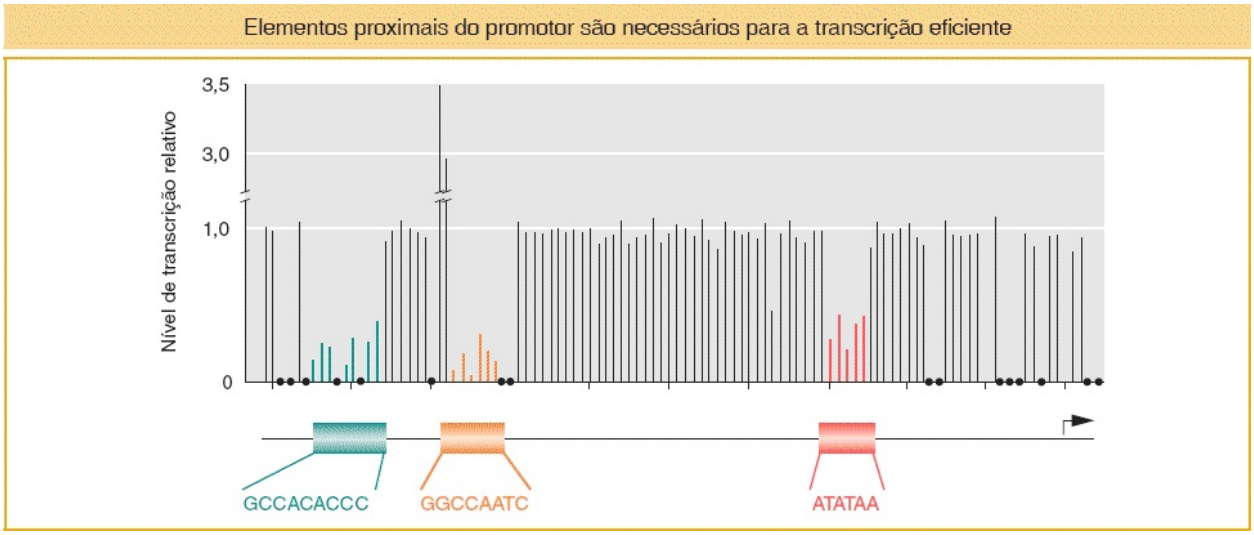


FIGURA 12.4 Mutações de ponto no promotor e nos elementos proximais do promotor impedem a transcrição do gene da β-globina. Mutações de ponto por toda a região do promotor foram analisadas em relação aos seus efeitos sobre as taxas de transcrição. A altura de cada linha representa o nível de transcrição em relação a um promotor do tipo selvagem ou elemento proximal do promotor (1,0). Apenas as substituições de bases que estão localizadas dentro dos três elementos demonstrados alteram o nível de transcrição. As

posições com *pontos pretos* não foram testadas. (Dados de T. Maniatis, S. Goodbourn e J. A. Fischer, "Regulation of Inducible and Tissue-Specific Gene Expression", *Science* 236, 1987, 1237.)

3. Um domínio que interage com proteínas ligadas a sequências reguladoras próximas no DNA, de tal modo que elas possam atuar de modo cooperativo para regular a transcrição.
4. Um domínio que influencia a condensação da cromatina, seja direta ou indiretamente.
5. Um domínio que atua como um sensor das condições fisiológicas dentro da célula.

Os mecanismos reguladores de genes eucarióticos foram descobertos por meio de abordagens bioquímicas e genéticas. As últimas sofreram avanços em particular por meio de estudos da levedura unicelular *Saccharomyces cerevisiae* (ver "Organismo-modelo", adiante). Esse organismo, que desempenhou um papel-chave em vinificação, produção de cervejas e panificação por muitos séculos, tem sido um passaporte para a compreensão de uma grande parte da biologia celular eucariótica. Diversas décadas de pesquisas produziram muitas percepções fundamentais sobre os princípios de como as proteínas reguladoras da transcrição eucariótica atuam e como os diferentes tipos celulares são gerados. Examinaremos dois sistemas reguladores de genes de levedura em detalhes: o primeiro se refere à via de utilização da galactose; o segundo é o controle do tipo reprodutivo.

12.2 Lições das leveduras | Sistema GAL

Para fazer uso da galactose extracelular, a levedura importa o açúcar e o converte em um tipo de glicose que pode ser metabolizada. Diversos genes – *GAL1*, *GAL2*, *GAL7* e *GAL10* – no genoma de levedura codificam enzimas que catalisam as etapas na via bioquímica que converte a galactose em glicose (Figura 12.5). Três genes adicionais – *GAL3*, *GAL4* e *GAL80* – codificam proteínas que regulam a expressão dos genes enzimáticos. Assim como no sistema *lac* de *E. coli*, a

abundância do açúcar determina o nível de expressão gênica na via bioquímica. Em células de levedura cultivadas em meios com ausência de galactose, os genes *GAL* são amplamente silenciados. Mas, na presença de galactose (e na ausência de glicose), os genes *GAL* são induzidos. Assim como em relação ao óperon *lac*, as análises genéticas e moleculares de mutantes têm sido uma chave para compreender como a expressão dos genes na via da galactose é controlada.

FIGURA 12.5 A galactose é convertida em glicose-1-fosfato em uma série de etapas. Essas etapas são catalisadas por enzimas (Gal1 e assim por diante) codificadas pelos genes estruturais *GAL1*, *GAL2*, *GAL7* e *GAL10*.

O regulador-chave da expressão gênica de *GAL* é a proteína Gal4, uma proteína de ligação ao DNA sequência-específica. A Gal4 talvez seja a proteína reguladora da transcrição mais bem-estudada em eucariotos. A dissecção detalhada da sua regulação e atividade tem sido uma fonte de diversas

percepções-chave sobre o controle da transcrição em eucariotos.

ORGANISMO-MODELO *Levedura*

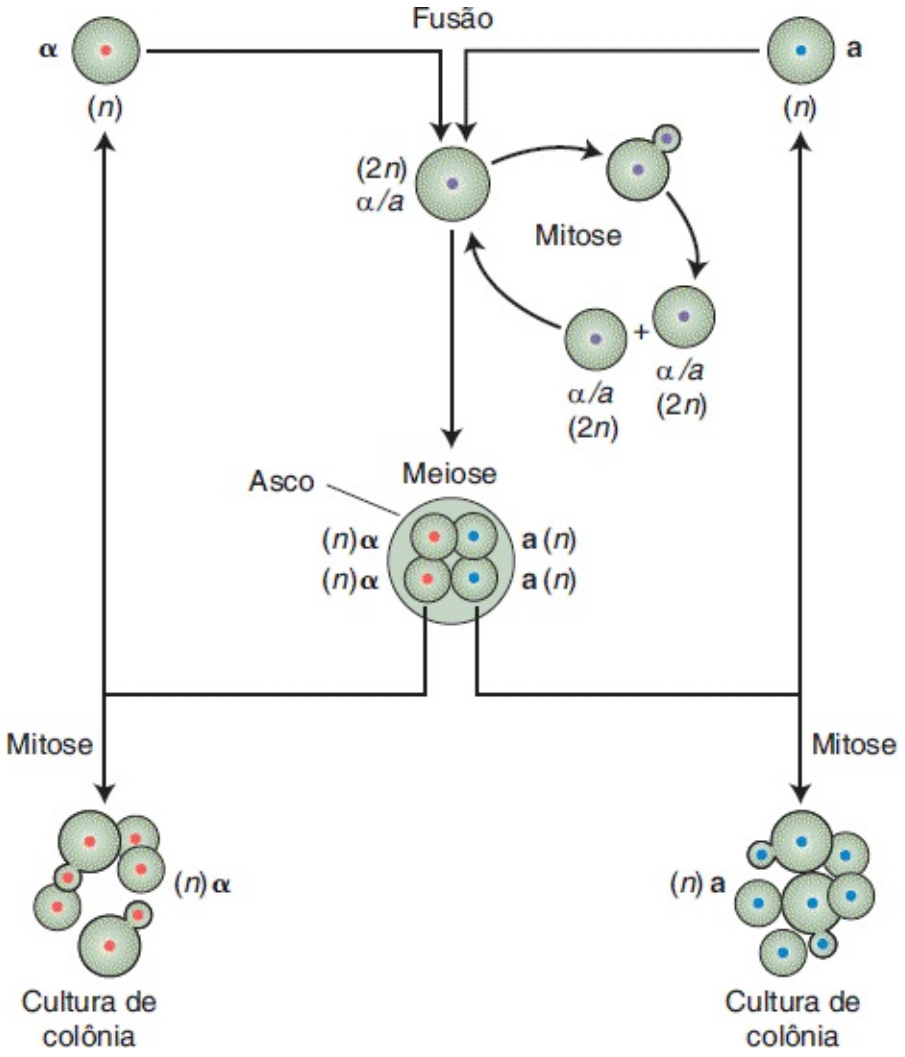
O *Saccharomyces cerevisiae* surgiu nos últimos anos como um importante sistema genético eucariótico. Os seres humanos têm cultivado leveduras há séculos, tendo em vista que são um componente essencial da cerveja, do pão e do vinho. A levedura apresenta muitas características que a tornam um organismo-modelo ideal. Como um eucarioto unicelular, ela pode ser cultivada em placas com ágar e, com o ciclo de vida da levedura de apenas 90 minutos, grandes quantidades podem ser cultivadas em meio líquido. Ela apresenta um genoma muito compacto, com aproximadamente apenas 12 megapares de bases de DNA (em comparação a quase 3.000 megapares de bases nos seres humanos), que contêm aproximadamente 6.000 genes distribuídos em 16 cromossomos. Foi o primeiro eucarioto a ter o seu genoma sequenciado.

O ciclo de vida da levedura a torna muito versátil para estudos laboratoriais. As células podem ser cultivadas como diploides ou haploides. Em ambos os casos, a célula-mãe produz um brotamento que contém uma célula-filha idêntica. As células diploides continuam a crescer por brotamento, ou são induzidas à meiose, que produz quatro esporos haploides mantidos unidos em um asco (também chamado tétrade). Os esporos haploides de tipo reprodutivo oposto (**a** ou α) se fundirão e formarão um diploide. Esporos do mesmo tipo reprodutivo continuarão a crescer por brotamento.

A levedura tem sido denominada a *E. coli* dos eucariotos, em virtude da facilidade da análise de mutantes direta e reversa. Para isolar mutantes por meio de uma abordagem genética direta, são induzidas mutações nas células haploides (com raios X, por exemplo) e triadas em placas em relação aos fenótipos mutantes. Esse procedimento normalmente é realizado primeiramente por meio do plaqueamento de células em um meio rico no qual todas as células crescem e por meio da realização de cópias, ou *plaqueamento em réplica*, das colônias dessa placa máster em placas de réplica que contêm meio seletivo ou condições de crescimento especiais. (Ver também o Capítulo 16.) Por

exemplo, mutantes sensíveis à temperatura crescerão na placa máster na temperatura permissiva, mas não na réplica em uma temperatura restritiva. A comparação das colônias nas placas máster e réplica revelará os mutantes sensíveis à temperatura. Com a utilização da genética reversa, os cientistas também conseguem substituir qualquer gene de levedura (de função conhecida ou desconhecida) por uma versão mutante (sintetizada em um tubo de ensaio) para compreender a natureza do produto gênico.

Micrografia eletrônica de células de levedura de brotamento. (SciMAT/ Science Souce.)



O ciclo de vida de *S. cerevisiae*. Os alelos nucleares *MAT***a** e *MAT*α determinam o tipo reprodutivo.

Gal4 regula múltiplos genes por meio de sequências de ativação upstream

Na presença de galactose, os níveis de expressão dos genes *GAL1*, *GAL2*, *GAL7* e *GAL10* são 1.000 vezes superiores, ou mais, do que na sua ausência. Em mutantes *GAL4*, entretanto, eles permanecem silenciosos. Cada um desses quatro genes apresenta dois ou mais sítios de ligação Gal4 localizados a alguma distância de 5′ (*upstream*) do seu promotor. Considere os genes *GAL10* e *GAL1*, que estão

adjacentes entre si e que são transcritos em sentidos opostos. Entre o sítio de início da transcrição de *GAL1* e o sítio de início da transcrição de *GAL10* está uma única região de 118 pb que contém quatro sítios de ligação Gal4 (Figura 12.6). Cada sítio de ligação Gal4 apresenta de 17 pares de bases de comprimento e está ligado por um dímero da proteína Gal4. Também existem dois sítios de ligação Gal4 *upstream* do gene *GAL2* e outros dois *upstream* do gene *GAL7*. Esses sítios de ligação são necessários para a ativação gênica *in vivo*. Se eles são deletados, os genes são silenciados, até mesmo na presença de galactose. Essas sequências reguladoras são acentuadores. A presença de acentuadores localizados a uma distância linear considerável do promotor de um gene eucariótico é típica. Tendo em vista que os acentuadores ativados por Gal4 estão localizados *upstream* (5′) dos genes que eles regulam, eles também são denominados **sequências de ativação *upstream* (UAS)**.

> CONCEITO-CHAVE A ligação de proteínas de ligação ao DNA sequência-específicas a regiões fora dos promotores dos genes-alvo é uma característica comum da regulação da transcrição eucariótica.

Proteína Gal4 apresenta domínios de ligação ao DNA e ativação separados

Após a ligação de Gal4 ao elemento UAS, como a expressão gênica é induzida? É necessário um domínio distinto da proteína Gal4, o **domínio de ativação**, para a atividade reguladora. Portanto, a proteína Gal4 apresenta no mínimo dois domínios: um para a ligação ao DNA e outro para a ativação da transcrição. Observou-se que uma organização modular semelhante é uma característica comum também de outros fatores de transcrição de ligação ao DNA.

A organização modular da proteína Gal4 foi demonstrada em uma série de experimentos. A estratégia era testar a ligação ao DNA e a ativação gênica de formas mutantes da proteína nas quais partes haviam sido deletadas ou fundidas a outras proteínas. Dessa maneira, os investigadores puderam determinar se uma

parte da proteína era necessária para uma função em particular. Para realizar esses estudos, os experimentadores necessitavam de um meio simples para analisar a expressão das enzimas codificadas pelos genes *GAL*.

A expressão dos genes *GAL* e outros alvos de fatores de transcrição é tipicamente monitorada por meio da utilização de um **gene repórter**, cujo nível de expressão é facilmente medido. Nos construtos de gene repórter, ele é ligado a sequências reguladoras que determinam a expressão do gene que está sendo investigado. A expressão do gene repórter reflete a atividade do elemento regulador que está sendo investigado. Com frequência, o gene repórter é o gene *lacZ* de *E. coli*. O *lacZ* é um gene repórter eficaz, pois os produtos de sua atividade são facilmente medidos. Outro gene repórter comum é o que codifica a proteína fluorescente verde (GFP) da água-viva. Conforme a sua denominação sugere, a concentração da proteína repórter é facilmente medida pela quantidade de luz que ela emite. Para investigar o controle da expressão do gene *GAL*, a região codificadora de um desses genes repórter e um promotor são posicionados *downstream* de um elemento UAS do gene *GAL*. A expressão do repórter é, então, uma leitura da atividade de Gal4 nas células (Figura 12.7 A).

Vejamos o que ocorre quando uma forma da proteína Gal4 com ausência do domínio de ativação é expressa em leveduras. Nesse caso, os sítios de ligação do elemento UAS estão ocupados, mas nenhuma transcrição é estimulada (Figura 12.7 B). O mesmo é verdadeiro quando outras proteínas reguladoras com ausência de domínios de ativação, tais como o repressor bacteriano LexA, são expressas nas células que contêm genes repórter com seus respectivos sítios de ligação. O resultado mais interessante é obtido quando o domínio de ativação da proteína Gal4 é inserido no domínio de ligação ao DNA da proteína LexA; a proteína híbrida agora ativa a transcrição dos sítios de ligação de LexA (Figura 12.7 D). Experimentos adicionais de “permuta de domínio” revelaram que a função de ativação da transcrição da proteína Gal4 reside em duas pequenas regiões com comprimento de aproximadamente 50 a 100 aminoácidos. Essas duas regiões formam um domínio de ativação em separado, que auxilia no recrutamento do maquinário da transcrição para o promotor, conforme veremos posteriormente nesta seção. Esse arranjo altamente modular dos domínios de regulação de

atividade é observado em muitos fatores de transcrição.

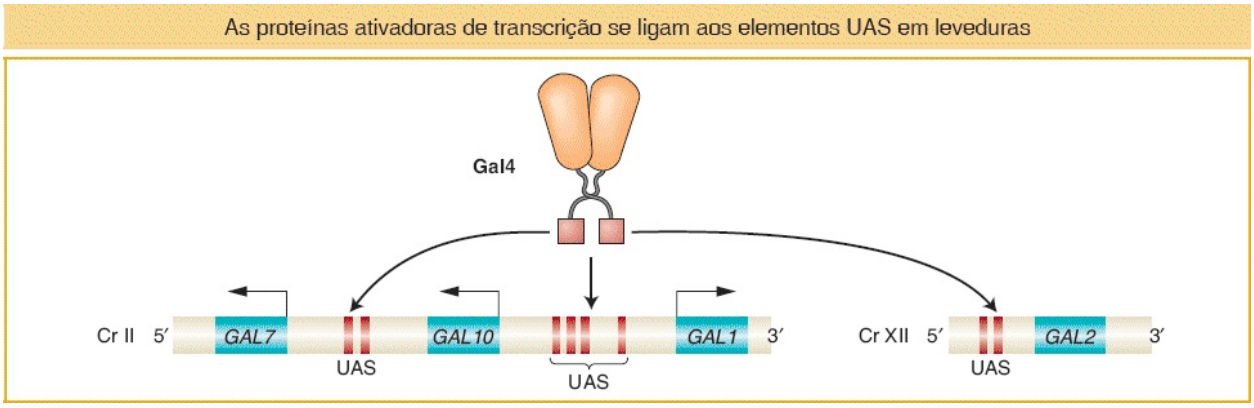


FIGURA 12.6 A proteína Gal4 ativa os genes-alvo por meio de elementos de sequência de ativação *upstream* (UAS). A proteína Gal4 apresenta dois domínios funcionais: um domínio de ligação ao DNA (*quadrado rosa*) e um domínio de ativação (*oval laranja*). A proteína se liga como um dímero a sequências específicas *upstream* dos promotores gênicos da via Gal. Alguns dos genes *GAL* são adjacentes (*GAL1*, *GAL10*), enquanto outros estão em cromossomos diferentes. O elemento UAS *GAL1* contém quatro sítios de ligação Gal4.

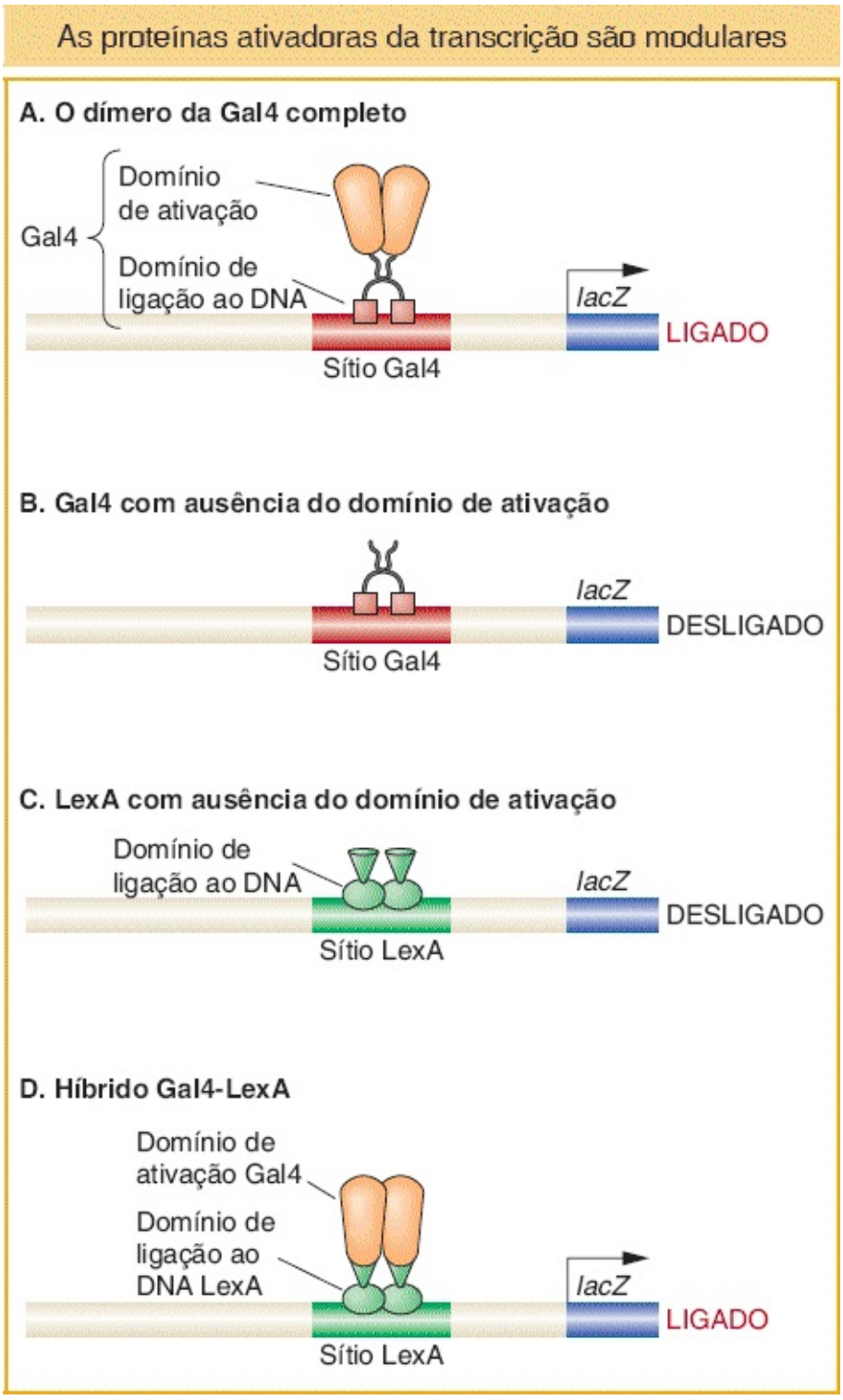


FIGURA 12.7 Proteínas ativadoras da transcrição apresentam múltiplos domínios separáveis. **A.** A proteína Gal4 apresenta dois domínios e forma um dímero. **B.** A remoção experimental do domínio de ativação demonstra que a ligação ao DNA não é suficiente para a ativação gênica. **C.** De modo semelhante, a proteína LexA bacteriana não consegue ativar a transcrição por si própria, mas, quando fundida ao domínio de ativação

Gal4 (**D**), consegue ativar a transcrição por meio dos sítios de ligação de LexA.

> CONCEITO-CHAVE Muitas proteínas eucarióticas reguladoras de transcrição são proteínas modulares, que apresentam domínios separados para a ligação ao DNA, ativação ou repressão e interação com outras proteínas.

Atividade da Gal4 é regulada fisiologicamente

Como a Gal4 se torna ativa na presença de galactose? As indicações-chave advêm da análise de mutações nos genes *GAL80* e *GAL3*. Em mutantes *GAL80*, os genes estruturais *GAL* estão ativos até mesmo na ausência de galactose. Esse resultado sugere que a função normal da proteína Gal80 é inibir, de algum modo, a expressão do gene *GAL*. Contrariamente, em mutantes *GAL3*, os genes estruturais *GAL* não estão ativos na presença de galactose, sugerindo que Gal3 normalmente promove a expressão dos genes *GAL*.

Análises bioquímicas extensivas revelaram que a proteína Gal80 se liga à proteína Gal4 com alta afinidade e inibe diretamente a atividade de Gal4. Especificamente, Gal80 se liga a uma região em um dos domínios de ativação de Gal4, bloqueando a sua capacidade de promover a transcrição dos genes-alvo. A proteína Gal80 é expressa continuamente, de modo que ela está sempre atuando para reprimir a transcrição dos genes estruturais *GAL*, exceto se for interrompida. O papel da proteína Gal3 é liberar os genes estruturais *GAL* da repressão pela Gal80 quando a galactose está presente.

Portanto, Gal3 é tanto um sensor quanto um indutor. Quando Gal3 se liga à galactose e ao ATP, ela é submetida a uma alteração alostérica que promove a ligação à Gal80, que por sua vez permite que Gal80 liberte Gal4, que então é capaz de interagir com outros fatores de transcrição e a RNA pol II para ativar a transcrição de seus genes-alvo. Portanto, Gal3, Gal80 e Gal4 são, todas, partes de uma mudança cujo estado é determinado pela presença ou pela ausência de galactose (Figura 12.8). Nessa mudança, a ligação do regulador transcricional ao DNA não é a etapa regulada fisiologicamente (como no caso no óperon *lac* e no

bacteriófago λ); em vez disso, a atividade do domínio de ativação é regulada.

> CONCEITO-CHAVE A atividade das proteínas reguladoras da transcrição eucariótica com frequência é controlada por interações com outras proteínas.

Funções da Gal4 na maior parte dos eucariotos

Além de sua ação em células de levedura, demonstrou-se que a Gal4 é capaz de ativar a transcrição em células de insetos, células humanas e de muitas outras espécies eucarióticas. Essa versatilidade sugere que o maquinário bioquímico e os mecanismos de ativação gênica são comuns a uma ampla variedade de eucariotos, que as características reveladas em leveduras em geral estão presentes em outros eucariotos e vice-versa. Além disso, em virtude da sua versatilidade, a Gal4 e seus elementos UAS se tornaram ferramentas favorecidas na análise genética para a manipulação da expressão e da função gênica em uma ampla variedade de sistemas-modelo.

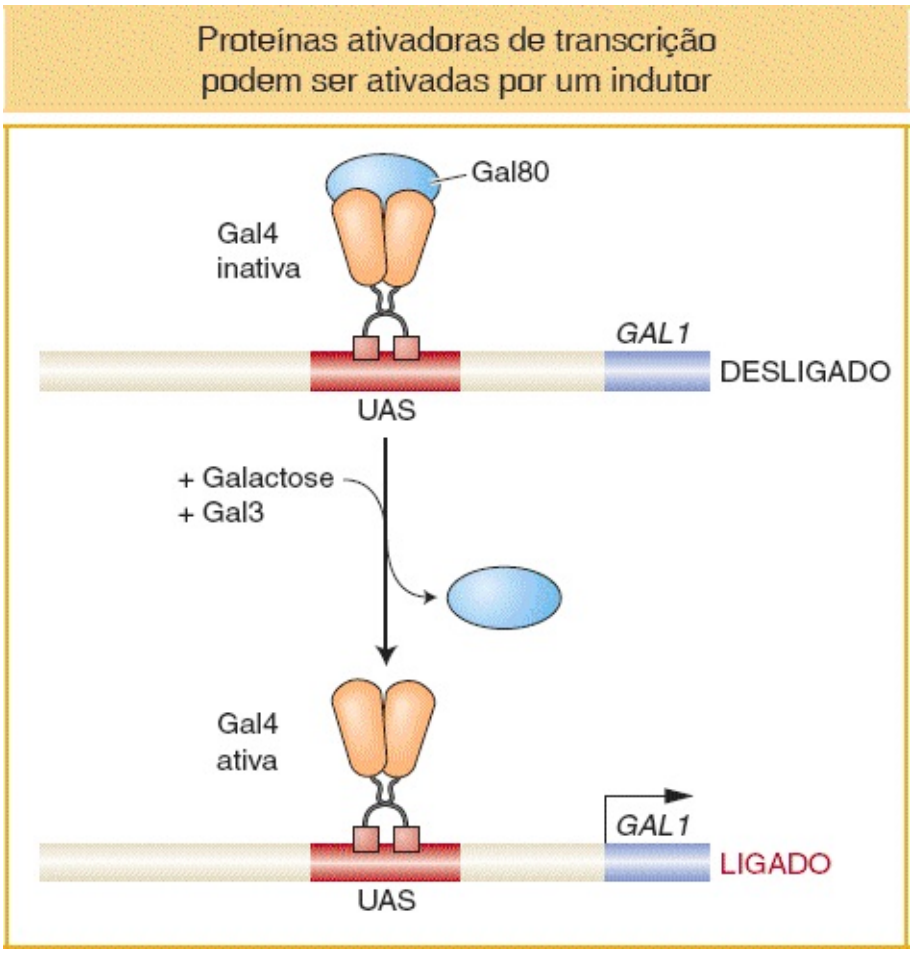


FIGURA 12.8 A atividade de Gal4 é regulada pela proteína Gal80. Na ausência de galactose, a proteína Gal4 é inativa, muito embora ela possa se ligar a sítios *upstream* do gene-alvo *GAL1*. A atividade de Gal4 é suprimida pela ligação da proteína Gal80 (*parte superior*). Na presença de galactose e da proteína Gal3, Gal80 sofre uma alteração na conformação e é liberada, possibilitando que o domínio de ativação Gal4 ative a transcrição do gene-alvo (*parte inferior*).

> CONCEITO-CHAVE A capacidade de Gal4, bem como de outros reguladores eucarióticos, de atuar em uma diversidade de eucariotos, indica que os eucariotos em geral apresentam o maquinário regulador da transcrição e mecanismos em comum.

Agora observaremos como os ativadores e outras proteínas reguladoras interagem com o maquinário de transcrição para controlar a expressão gênica.

Ativadores recrutam o maquinário de transcrição

Em bactérias, ativadores comumente estimulam a transcrição por meio da interação diretamente com o DNA e com a RNA polimerase. Em eucariotos, os ativadores em geral atuam indiretamente. Os ativadores eucarióticos recrutam a RNA polimerase II para promotores gênicos por meio de dois mecanismos importantes. Primeiramente, os ativadores podem interagir com subunidades dos complexos proteicos que atuam no início da transcrição e em seguida os recrutam para o promotor. Em segundo lugar, os ativadores podem recrutar proteínas que modificam a estrutura da cromatina, possibilitando à RNA polimerase II e a outras proteínas o acesso ao DNA. Muitos ativadores, incluindo Gal4, apresentam ambas as atividades. Primeiramente, examinaremos o recrutamento de partes do complexo de iniciação transcricional.

Lembre, do Capítulo 8, que o maquinário da transcrição eucariótica contém muitas proteínas que são parte de diversos subcomplexos no aparato de transcrição, o que é montado nos promotores gênicos. Um subcomplexo, o fator de transcrição IID (TFIID), liga-se ao TATA boxe de promotores eucarióticos por meio da proteína de ligação a TATA (TBP; ver Figura 8.12). Um modo pelo qual a Gal4 atua para ativar a expressão gênica é pela ligação à TBP em um sítio no seu domínio de ativação. Por meio dessa interação de ligação, recruta o complexo TFIID e, por sua vez, a RNA polimerase II para o promotor (Figura 12.9). A força dessa interação entre Gal4 e TBP está bem correlacionada com a potência de Gal4 como um ativador.

Um segundo modo pelo qual a Gal4 atua para ativar a expressão gênica é pela interação com o **complexo mediador**, um grande complexo multiproteico que, por sua vez, interage diretamente com a RNA polimerase II para recrutá-la para os promotores gênicos. O complexo mediador é um exemplo de **coativador**, termo aplicado a uma proteína ou um complexo proteico que facilita a ativação gênica por meio de um fator de transcrição, mas que, por si próprio, não é parte do maquinário de transcrição, nem é uma proteína de ligação ao DNA.

A capacidade dos fatores de transcrição de se ligarem às sequências do DNA *upstream* e interagirem com proteínas que se ligam direta ou indiretamente aos

promotores ajuda a explicar como a transcrição pode ser estimulada por sequências reguladoras mais distantes (ver Figura 12.9).

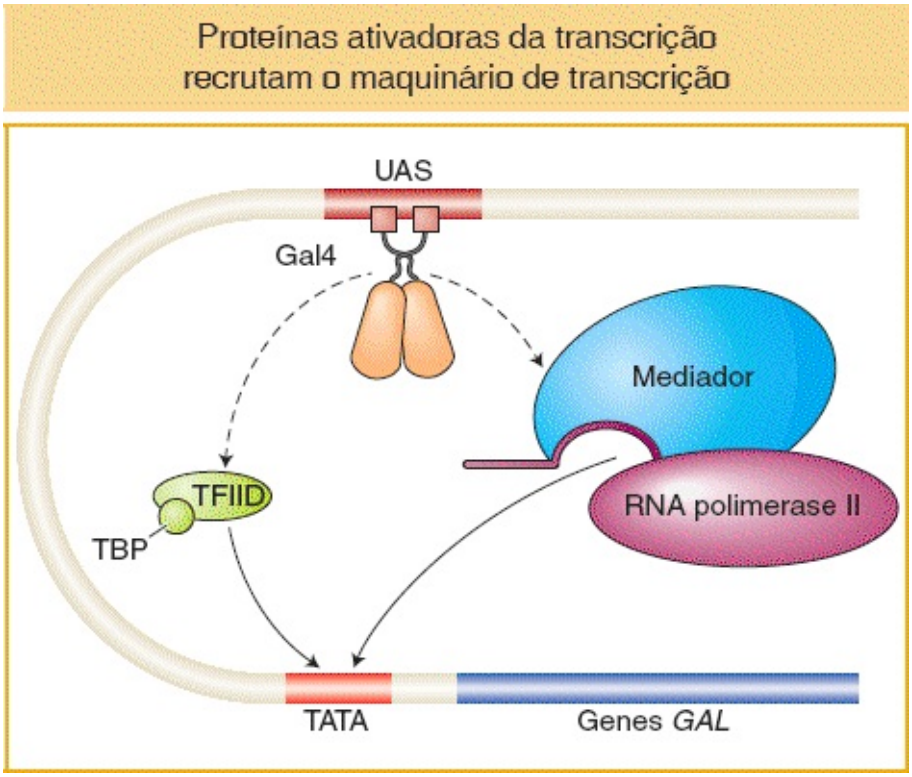


FIGURA 12.9 Gal4 recruta o maquinário de transcrição. A proteína Gal4 e muitos outros ativadores de transcrição se ligam a múltiplos complexos proteicos, incluindo o TFIID e os complexos mediadores demonstrados aqui (*setas tracejadas*), que recrutam a RNA polimerase II para os promotores gênicos. As interações facilitam a ativação gênica pelos sítios de ligação que estão distantes dos promotores gênicos.

> CONCEITO-CHAVE Os ativadores da transcrição eucariótica com frequência atuam por meio do recrutamento de partes do maquinário de transcrição para os promotores gênicos.

Controle do tipo reprodutivo de leveduras | Interações combinatórias

Até agora, enfocamos, neste capítulo, a regulação de genes únicos ou de alguns genes em uma via. Em organismos multicelulares, tipos celulares distintos diferem na expressão de centenas de genes. A expressão ou a repressão de grupos de genes, portanto, deve ser coordenada na formação de tipos celulares em

particular. Um dos exemplos mais bem-compreendidos de regulação do tipo celular em eucariotos é a regulação do tipo reprodutivo em leveduras. Esse sistema regulador foi dissecado por uma combinação de genética, biologia molecular e bioquímica. O tipo reprodutivo atua como um excelente modelo para a compreensão da lógica da regulação gênica em animais multicelulares.

Saccharomyces cerevisiae pode existir em qualquer um de três tipos celulares diferentes, conhecidos como **a**, α e **a**/α. Os dois tipos celulares **a** e α são haploides e contêm apenas uma cópia de cada cromossomo. A célula **a**/α é diploide e contém duas cópias de cada cromossomo. Embora os dois tipos de células haploides não possam ser distinguidos por seu aspecto ao microscópio, eles podem ser diferenciados por meio de uma diversidade de características celulares específicas, principalmente do seu tipo reprodutivo (ver "Organismo-modelo", anteriormente). Uma célula α cruza apenas com uma célula **a** e uma célula **a** cruza apenas com uma célula α. Uma célula α secreta um oligopeptídio *feromônio*, ou hormônio sexual, denominado fator α, que paralisa as células **a** no ciclo celular. De modo semelhante, uma célula **a** secreta um feromônio, denominado fator **a**, que paralisa as células α. A parada celular de ambos os participantes é necessária para o sucesso reprodutivo. A célula diploide **a**/α não realiza cruzamento, é maior do que as células α e **a** e não responde aos hormônios reprodutivos.

A análise genética de mutantes defeituosos no cruzamento demonstrou que o tipo celular é controlado por um único *locus* genético, o *locus* do tipo reprodutivo, *MAT*. Existem dois alelos do *locus MAT*: as células haploides **a** apresentam o alelo *MATa* e as células haploides α apresentam o alelo *MAT*α. A célula diploide **a**/α apresenta ambos os alelos. Embora o tipo reprodutivo esteja sob o controle genético, determinadas linhagens alteram o seu tipo reprodutivo, por vezes tão frequentemente quanto a cada divisão celular. Examinaremos a base da alteração posteriormente neste capítulo, mas, primeiramente, vejamos como cada tipo celular expressa o grupo de genes correto. Veremos que diferentes combinações de proteínas de ligação ao DNA regulam a expressão de conjuntos de genes específicos para tipos celulares diferentes.

Como o *locus MAT* controla o tipo celular? Análises genéticas de mutantes que

não podem cruzar identificaram uma quantidade de genes estruturais que estão separados do *locus MAT*, mas cujos produtos proteicos são necessários para o cruzamento. Um grupo de genes estruturais é expresso apenas no tipo celular α (genes específicos de α) e outro grupo é expresso apenas no tipo celular **a** (genes **a**-específicos). O *locus MAT* controla quais desses grupos de genes estruturais são expressos em cada tipo celular. O alelo *MATa* causa a expressão dos genes estruturais da célula do tipo **a**, enquanto o alelo *MAT*α causa a expressão dos genes estruturais da célula do tipo α. Esses dois alelos ativam diferentes grupos de genes, tendo em vista que codificam diferentes proteínas reguladoras. Além disso, uma proteína reguladora não codificada pelo *locus MAT*, denominada MCM1, desempenha um papel-chave na regulação do tipo celular.

O caso mais simples é o tipo celular **a** (Figura 12.10 A). O *locus MATa* codifica uma única proteína reguladora, a1. Entretanto, a1 não apresenta efeito em células haploides, apenas em células diploides. Em uma célula haploide **a**, a proteína reguladora Mcm1 ativa a expressão dos genes estruturais necessários para uma célula **a**, ligando-se às sequências reguladoras em promotores gênicos **a**-específicos.

Em uma célula α, os genes estruturais α-específicos devem ser transcritos, mas, além disso, deve-se evitar que a proteína MCM1 ative genes **a**-específicos. A sequência do DNA do alelo *MAT*α codifica duas proteínas, α1 e α2, que são produzidas por unidades de transcrição em separado. Essas duas proteínas apresentam papéis reguladores diferentes na célula, conforme pode ser demonstrado por meio da análise das suas propriedades de ligação do DNA *in vitro* (Figura 12.10 B). A proteína α1 é um ativador de expressão gênica α-específico. Ela se liga em combinação com a proteína MCM1 a uma sequência discreta de DNA que controla diversos genes α-específicos. A proteína α2 reprime a transcrição de genes **a**-específicos. Ela se liga como um dímero, com MCM1, a sítios nas sequências de DNA localizadas *upstream* de um grupo de genes **a**-específicos e atua como um repressor.

Em uma célula diploide de levedura, as proteínas reguladoras codificadas por cada *locus MAT* são expressas (Figura 12.10 C). Qual é o resultado? Todos os genes estruturais envolvidos no cruzamento celular são desligados, pois são um

conjunto de genes em separado, denominados haploides específicos, que é expresso em células haploides, mas não em células diploides. Como isso ocorre? A proteína a1 codificada por *MATa* tem, finalmente, uma função a desempenhar. A proteína a1 consegue se ligar a algumas das proteínas α2 presentes e alterar sua especificidade de ligação, de modo que o complexo de a1-α2 não se liga a genes **a**-específicos. Em vez disso, o complexo de a1-α2 se liga a uma sequência diferente, localizada *upstream* dos genes haploide-específicos. Então, em células diploides, a proteína α2 se apresenta em duas formas: (1) como um complexo α2-MCM1 que reprime genes **a**-específicos e (2) em um complexo com a proteína a1 que reprime genes específicos de haploides. Além disso, o complexo de a1-α2 também reprime a expressão do gene α1, que, assim, deixa de estar presente para ligar os genes α-específicos. Os diferentes parceiros de ligação determinam quais sequências de DNA específicas são ligadas e quais genes são regulados por cada um dos complexos que contêm α2. A regulação de diferentes conjuntos de genes-alvo por meio da associação do mesmo fator de transcrição com diferentes parceiros de ligação desempenha um papel importante na geração de diferentes padrões de expressão gênica em diferentes tipos celulares dentro de eucariotos multicelulares.

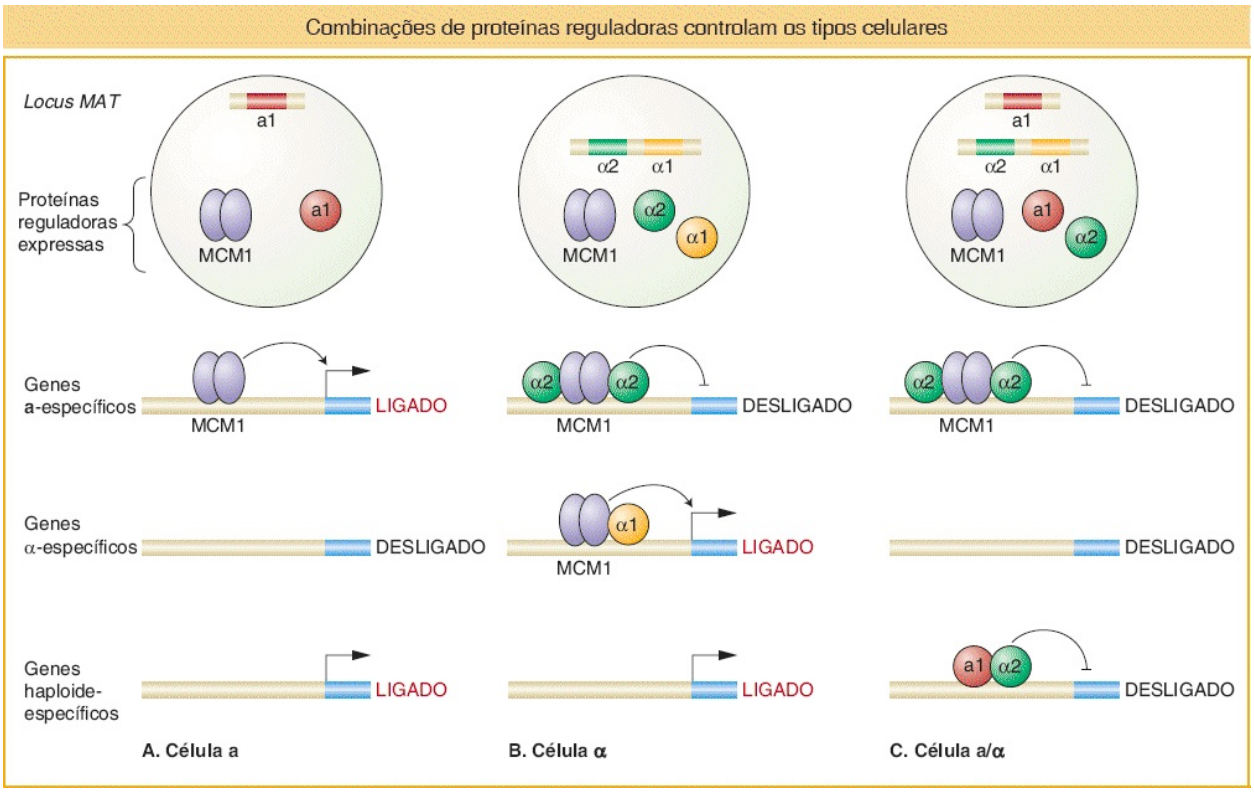


FIGURA 12.10 Controle da expressão gênica específica do tipo celular em leveduras. Os três tipos celulares de *S. cerevisiae* são determinados pelas proteínas reguladoras a1, α1 e α2, que regulam diferentes subconjuntos de genes-alvo. A proteína MCM1 atua em todos os três tipos celulares e interage com α1 e α2.

> CONCEITO-CHAVE Em leveduras e em eucariotos multicelulares, os padrões de expressão gênica específicos do tipo celular são regulados por combinações de fatores de transcrição que interagem.

12.3 Cromatina dinâmica

Um segundo mecanismo que influencia a transcrição gênica em eucariotos modifica a estrutura da cromatina local ao redor das sequências gênicas reguladoras. Para compreender totalmente como esse mecanismo atua, primeiramente precisamos compreender a estrutura da cromatina e em seguida considerar como ela pode ser alterada e como essas alterações afetam a expressão gênica.

O recrutamento do maquinário de transcrição por ativadores pode aparentar ser um pouco semelhante em eucariotos e bactérias, com a principal diferença sendo o número de proteínas que interagem no maquinário de transcrição. De fato, há duas décadas, muitos biólogos imaginavam a regulação eucariótica simplesmente como uma versão bioquimicamente mais complicada daquela que havia sido descoberta em bactérias. Entretanto, essa visão foi alterada dramaticamente na medida em que os biólogos consideraram o efeito da organização do DNA genômico em eucariotos.

Em comparação ao DNA eucariótico, o DNA bacteriano é relativamente “nu”, o que o torna prontamente acessível para a RNA polimerase. Contrariamente, os cromossomos eucarióticos estão compactados na cromatina, que é composta por DNA e proteínas (principalmente histonas). A unidade básica da cromatina é o nucleossomo, que contém aproximadamente 150 pb de DNA envolto 1,7 vez ao redor de um cerne de proteínas histonas (Figura 12.11). O cerne do nucleossomo contém oito histonas, duas subunidades de cada uma das quatro histonas: histonas 2A, 2B, 3 e 4 (denominadas H2A, H2B, H3 e H4), organizadas como dois dímeros de H2A e H2B e um tetrâmero de H3 e H4. Circundando o cerne do nucleossomo encontra-se uma histona ligante, H1, que consegue compactar os nucleossomos em estruturas de ordem superior que condensam ainda mais o DNA.

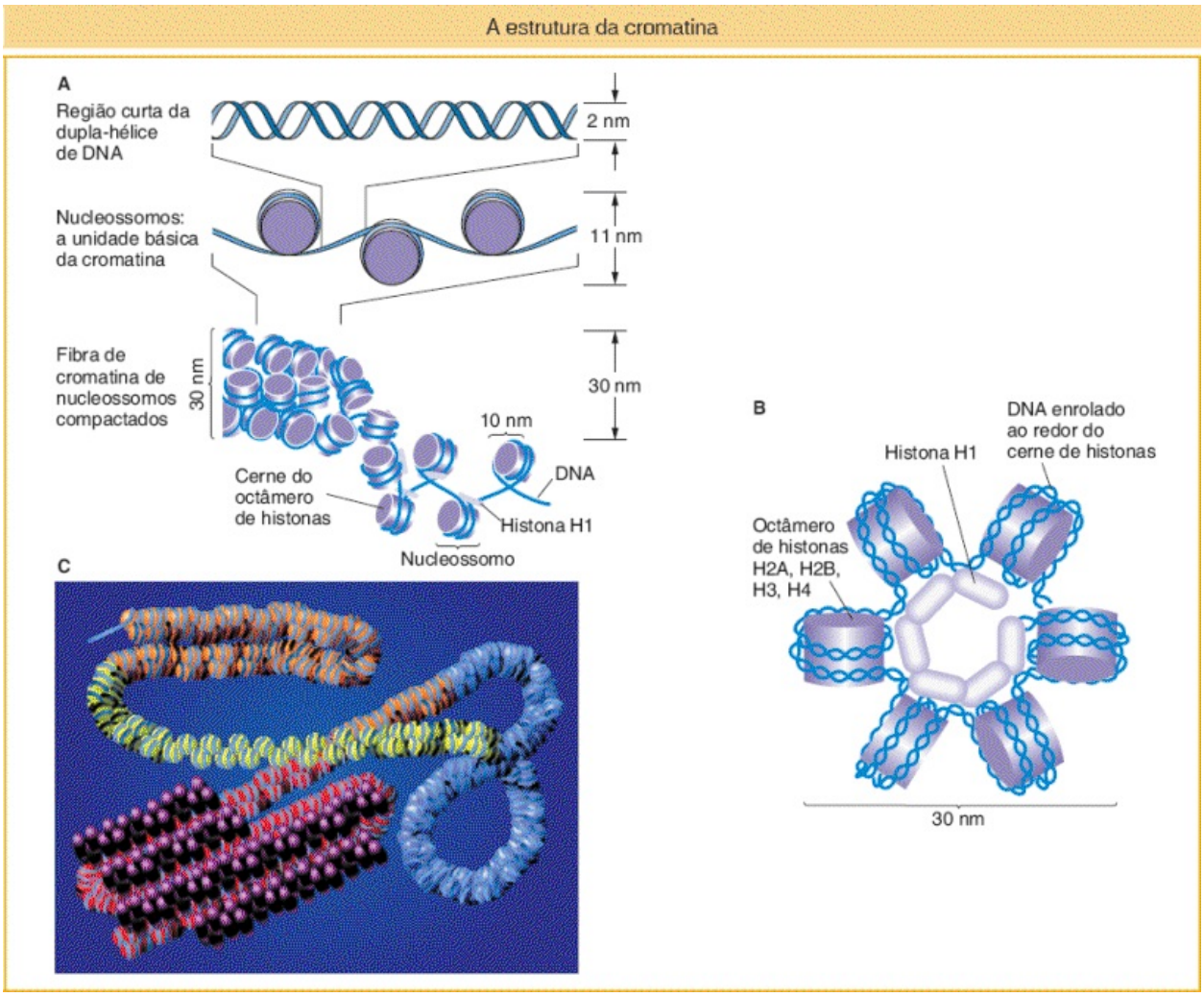


FIGURA 12.11 A. O nucleossomo na cromatina descondensada e condensada. **B.** Vista final da cadeia enrolada de nucleossomos. **C.** A estrutura da cromatina varia ao longo do comprimento de um cromossomo. A cromatina menos condensada (eucromatina) está demonstrada em *amarelo*, as regiões de condensação intermediária estão em *laranja* e *azul*, e a heterocromatina revestida por proteínas especiais (*roxo*) está em *vermelho*. (*C*. De P. J. Horn e C. L. Peterson, "Chromatin Higher Order Folding: Wrapping Up Transcription", *Science* 297, 2002, 1827, Fig. 3. Copyright 2002, AAAS.)

A compactação do DNA eucariótico em cromatina significa que uma grande parte do DNA não está prontamente acessível às proteínas reguladoras e ao aparato de transcrição. Portanto, enquanto os genes procarióticos em geral estão acessíveis e "ligados", exceto se reprimidos, os genes eucarióticos estão inacessíveis e "desligados", exceto se ativados. Portanto, a modificação da estrutura da cromatina é uma característica distintiva de muitos processos

eucarióticos, incluindo a regulação gênica (discutida neste capítulo), a replicação do DNA (Capítulo 7) e o reparo do DNA (Capítulo 16). Existem três mecanismos importantes para alterar a estrutura da cromatina, que serão discutidos de modo aprofundado nesta seção:

1. Movimentação dos nucleossomos ao longo do DNA, também denominada **remodelagem da cromatina**.
2. **Modificação das histonas** no cerne do nucleossomo.
3. Substituição das histonas comuns em um nucleossomo por **variantes de histonas**.

Proteínas remodeladoras da cromatina e ativação gênica

Um modo de alterar a estrutura da cromatina pode ser simplesmente mover o octâmero de histonas ao longo do DNA. Na década de 1980, foram desenvolvidas técnicas bioquímicas que possibilitaram que os pesquisadores determinassem a posição dos nucleossomos em e ao redor de genes específicos. Nesses estudos, a cromatina foi isolada de tecidos ou células nos quais um gene estava ligado e comparada à cromatina do tecido no qual o mesmo gene estava desligado. O resultado em relação à maior parte dos genes analisados foi que as posições dos nucleossomos estavam alteradas, especialmente nas regiões reguladoras dos genes. Portanto, as regiões do DNA que estão enroladas nos nucleossomos podem ser alteradas: as posições do nucleossomo podem ser alteradas no DNA de célula para célula e ao longo do ciclo de vida de um organismo. A transcrição é reprimida quando o promotor e as sequências flanqueadoras estão enrolados em um nucleossomo, o que evita o início da transcrição pela RNA pol II. Portanto, a ativação da transcrição requer que os nucleossomos sejam deslocados para longe do promotor. Contrariamente, quando a repressão gênica é necessária, os nucleossomos são alterados para uma posição que evita a transcrição. A alteração da posição do nucleossomo é denominada remodelagem da cromatina. Sabidamente a remodelagem da cromatina é uma parte integral da expressão gênica eucariótica e estão sendo feitos grandes avanços na determinação do(s) mecanismo(s) subjacentes e das proteínas reguladoras participantes. Aqui,

novamente, os estudos genéticos em leveduras foram essenciais.

Duas triagens genéticas em leveduras na busca por mutantes em processos aparentemente não relacionados levaram à descoberta do mesmo gene, cujo produto desempenha um papel-chave na remodelagem da cromatina. Em ambos os casos, as células de leveduras foram tratadas com agentes que causariam mutações. Em uma triagem, essas células de leveduras mutagenizadas foram triadas na busca por células que não cresciam adequadamente em sacarose (mutantes *n*ão *f*ermentadores de açúcar [*sugar*, em inglês], *snf*). Em outra triagem, células de leveduras mutagenizadas foram triadas na busca por mutantes que fossem defeituosos em alterar seu tipo reprodutivo (mutantes de alteração, *swi*; ver Seção 12.5). Muitos mutantes em relação a diferentes *loci* foram recuperados em cada triagem, mas observou-se que um gene mutante causa ambos os fenótipos. Os mutantes no assim denominado *locus swi2/snf2* (*switch-sniff*) não conseguem utilizar a sacarose efetivamente nem alterar o tipo reprodutivo.

Qual era a conexão entre a capacidade de utilizar o açúcar e a capacidade de alterar os tipos reprodutivos? A proteína Swi2-Snf2 foi purificada e descobriu-se que é parte de um grande complexo multissubunitário denominado complexo de SWI-SNF, que consegue reposicionar os nucleossomos em uma análise em tubo de ensaio, se o ATP for fornecido como uma fonte de energia (Figura 12.12). Em algumas situações, o complexo multissubunitário SWI-SNF ativa a transcrição por meio da movimentação dos nucleossomos que estão recobrindo as sequências TATA. Desse modo, o complexo facilita a ligação da RNA polimerase II. O complexo de SWI-SNF, portanto, é um coativador.

A Gal4 também se liga ao complexo de remodelagem da cromatina SWI-SNF e o recruta para promotores ativados. Linhagens de leveduras que contêm um complexo SWI-SNF defeituoso demonstram um nível reduzido de atividade de Gal4. Por que um ativador pode utilizar múltiplos mecanismos de ativação? Existem no mínimo dois motivos compreendidos no momento. O primeiro é que os promotores-alvo podem se tornar menos acessíveis em determinados estágios do ciclo celular ou em determinados tipos celulares (em eucariotos multicelulares). Por exemplo, os genes estão menos acessíveis durante a mitose, quando a cromatina está mais condensada. Naquele estágio, a Gal4 deve recrutar

o complexo de remodelagem da cromatina, enquanto em outras ocasiões, pode não ser necessário utilizar o complexo.

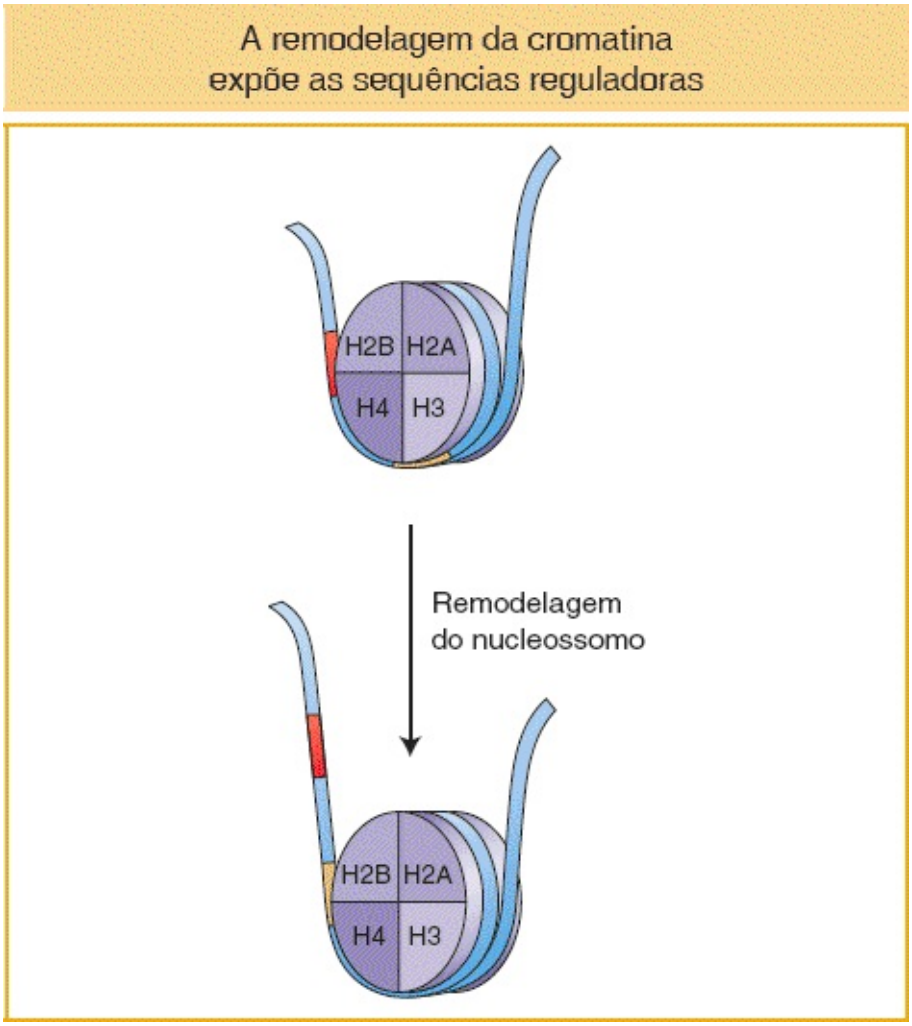


FIGURA 12.12 O octâmero de histonas desliza em resposta à atividade de remodelagem da cromatina (tal como aquela do complexo SWI-SNF), neste caso expondo o DNA marcado em *vermelho*. (Ver Figura 12.19 para os detalhes sobre como o SWI-SNF é recrutado para uma região do DNA em particular.)

Um segundo motivo é que muitos fatores de transcrição atuam em combinações para controlar a expressão gênica de modo sinérgico. Veremos brevemente que essa sinergia combinatória é resultante do fato de que os complexos de remodelagem da cromatina e o maquinário de transcrição são recrutados de modo mais eficiente quando diversos fatores de transcrição atuam em conjunto.

> CONCEITO-CHAVE A cromatina pode ser dinâmica; os nucleossomos não necessariamente estão em posições fixas no cromossomo. A remodelagem da cromatina altera a densidade ou a posição do nucleossomo e é uma parte integrante da regulação gênica eucariótica.

Modificação das histonas

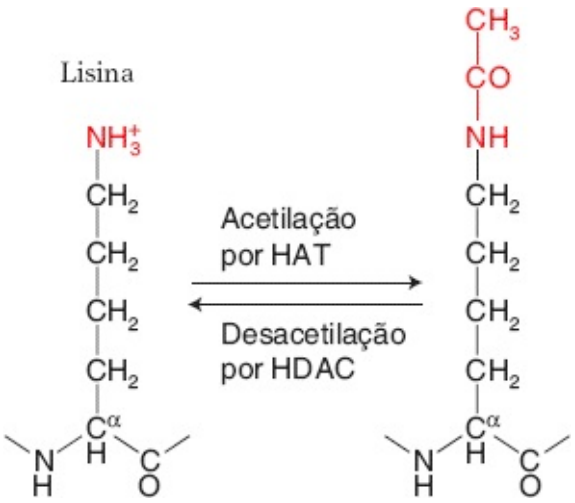
Vejamos o nucleossomo mais de perto para verificar se alguma parte dessa estrutura poderia carregar a informação necessária para influenciar a posição do nucleossomo, a densidade do nucleossomo ou ambas.

Conforme já declarado, a maior parte dos nucleossomos é composta por um octâmero de histonas composto por dois dímeros de H2A e H2B e um tetrâmero de H3 e H4. Sabidamente as histonas são as proteínas mais conservadas na natureza; ou seja, as histonas são quase idênticas em todos os organismos eucarióticos, desde leveduras até plantas e até animais. No passado, essa conservação contribuiu para a visão de que as histonas não poderiam participar em nada mais complicado do que a compactação do DNA para o encaixe no núcleo. Entretanto, relembre que o DNA, apenas com suas quatro bases, chegou a ser considerado uma molécula muito simples para conter o mapa de todos os organismos sobre a Terra.

A Figura 12.13 A demonstra um modelo de estrutura de nucleossomo que representa contribuições de muitos estudos. Digno de nota é que as proteínas histonas estão organizadas no cerne do octâmero, com algumas de suas extremidades aminoterminais realizando contatos eletrostáticos com o esqueleto de fosfato do DNA circundante. Essas extremidades em protrusão são denominadas **caudas de histonas**. Desde o início da década de 1960, sabe-se que resíduos de aminoácidos básicos específicos (lisina e arginina) nas caudas de histonas podem ser modificados de modo covalente por meio da ligação de grupos acetil e metil (Figura 12.13 B). Essas reações, que ocorrem após a proteína histona ter sido traduzida e até mesmo após a histona ter sido incorporada ao nucleossomo, são denominadas **modificações pós-tradução (PTM)**.

Sabe-se que existem no mínimo 150 modificações diferentes de histonas que utilizam uma ampla variedade de moléculas além dos grupos acetil e metil já mencionados, incluindo fosforilação, ubiquitinação e ribosilação de ADP. Diz-se que a modificação covalente das caudas de histonas contribui para um **código de histonas**. Cientistas cunharam a expressão "código de histonas" em virtude de a modificação covalente das caudas de histonas ser reminiscente do código genético. Em relação ao código de histonas, a informação é armazenada nos padrões de modificação de histonas, em vez de na sequência de nucleotídios. Com mais de 150 modificações de histonas conhecidas, existe um número enorme de possíveis padrões e os cientistas estão apenas começando a decifrar os seus efeitos sobre a estrutura da cromatina e a regulação da transcrição. Para aumentar essa complexidade, o código provavelmente não é interpretado precisamente do mesmo modo em todos os organismos. O papel da acetilação e da metilação das histonas na expressão gênica está descrito a seguir.

Acetilação e desacetilação de histonas e expressão gênica. A reação de acetilação é uma das modificações das histonas mais bem-caracterizadas:



Observe que a reação é reversível, o que significa que grupos acetil podem ser adicionados pela enzima histona acetiltransferase (HAT) e removidos pela enzima **histona desacetilase (HDAC)** do mesmo resíduo de histona. Por enquanto, veremos como a acetilação e a desacetilação de aminoácidos de histonas

influenciam a estrutura da cromatina e a expressão gênica.

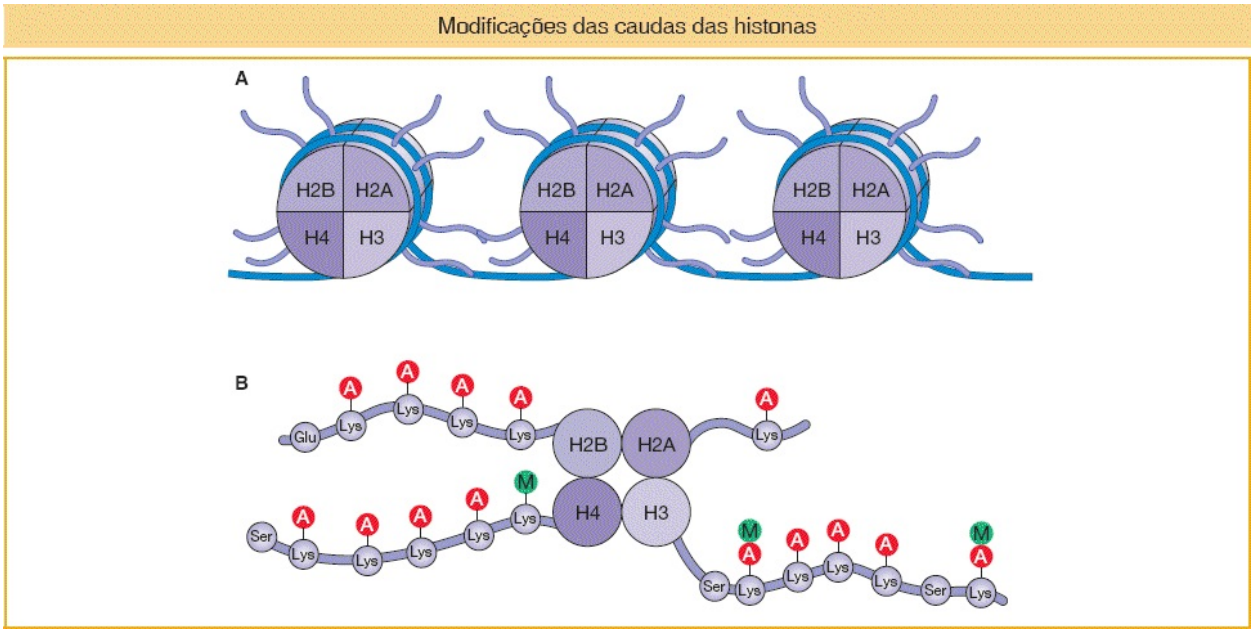


FIGURA 12.13 A. As caudas das histonas se projetam a partir do cerne do nucleossomo (*vermelho*). **B.** Estão demonstrados exemplos de modificações nas caudas das histonas. Os círculos com A representam acetilação, enquanto os círculos com M representam metilação. Ver o texto para os detalhes.

Há anos tem ocorrido o acúmulo de evidências de que as histonas associadas aos nucleossomos de genes ativos são ricas em grupos acetil (diz-se que são **hiperacetiladas**), enquanto aquelas associadas aos genes inativos são subacetiladas (**hipoacetiladas**). A enzima HAT comprovou ser muito difícil de isolar. Quando ela finalmente foi isolada e sua sequência proteica foi deduzida, observou-se que é *ortóloga* de um ativador de transcrição de levedura denominado GCN5 (o que significa que foi codificada pelo mesmo gene em um organismo diferente). Portanto, a conclusão foi que GCN5 é uma histona acetiltransferase. Ele se liga ao DNA nas regiões reguladoras de alguns genes e ativa a transcrição por meio da acetilação de histonas próximas. Atualmente compreende-se que diversos complexos proteicos que são recrutados por ativadores de transcrição possuem atividade de HAT.

Como a acetilação das histonas altera a estrutura da cromatina e, no processo, facilita as alterações na expressão gênica? A adição de grupos acetil aos resíduos

de histonas neutraliza a carga positiva dos resíduos de lisina e reduz a interação das caudas de histonas com a espinha dorsal do DNA carregada negativamente. Isso resulta em uma cromatina mais aberta na medida em que as interações eletrostáticas dos nucleossomos adjacentes e dos nucleossomos com o DNA adjacente são reduzidas (Figura 12.14). Além disso, a acetilação de histonas, em conjunto com outras modificações de histonas, influencia a ligação das proteínas reguladoras ao DNA. Uma proteína reguladora ligada pode participar em uma de diversas funções que direta ou indiretamente aumentam a frequência de início da transcrição.

Assim como outras modificações de histonas, a acetilação é reversível, e HDAC desempenham papéis-chave na repressão gênica. Por exemplo, na presença de galactose e glicose, a ativação de genes *GAL* é evitada pela proteína Mig1. A Mig1 é um repressor de ligação a sequências específicas de DNA que se liga a um sítio entre o elemento UAS e o promotor do gene *GAL1* (Figura 12.15). Mig1 recruta um complexo proteico denominado Tup1, que contém uma histona desacetilase e que reprime a transcrição gênica. O complexo Tup1 é um exemplo de **correpressor**, que facilita a repressão gênica, mas que por si próprio não é um repressor de ligação ao DNA. O complexo Tup1 também é recrutado por outros repressores de leveduras, tais como MATα2 (ver anteriormente) e correspondentes desse complexo são observados em todos os eucariotos.

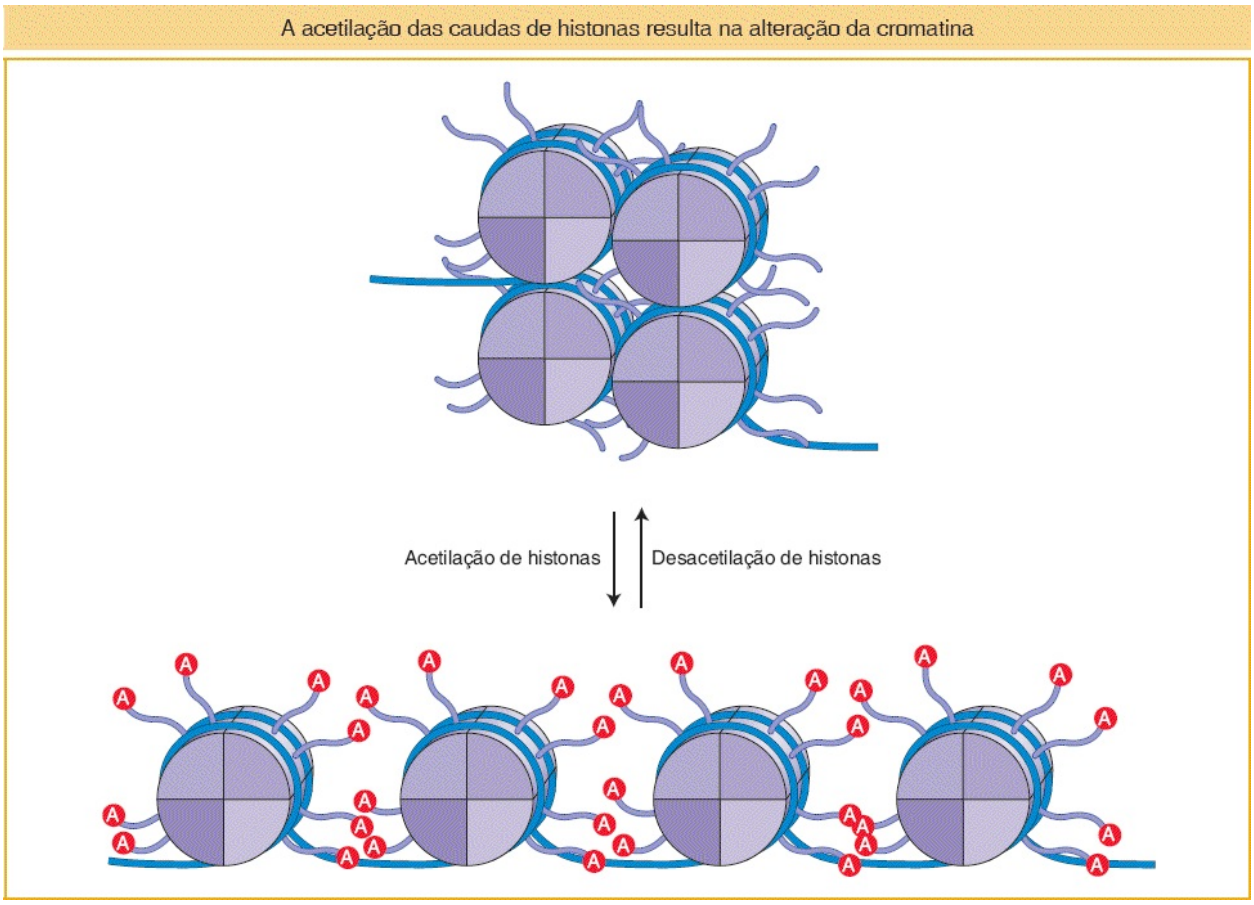


FIGURA 12.14 A acetilação dos aminoácidos lisina nas caudas de histonas abre a cromatina, expondo o DNA à atividade das proteínas que regulam a transcrição.

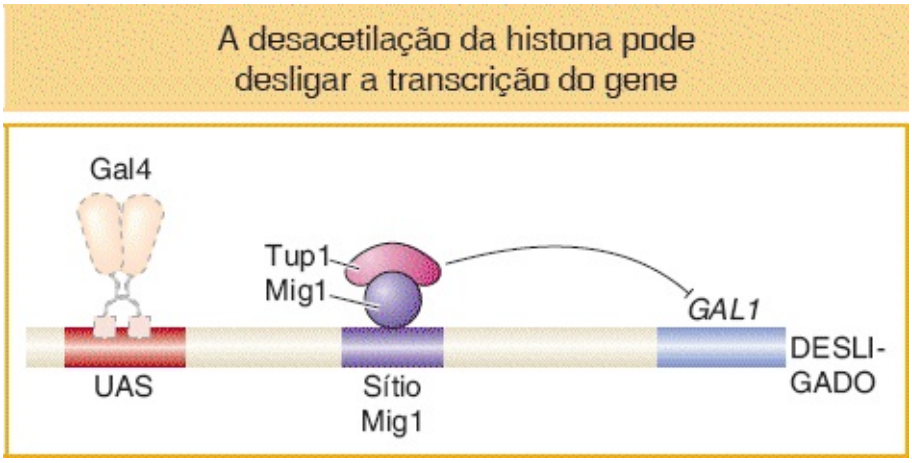
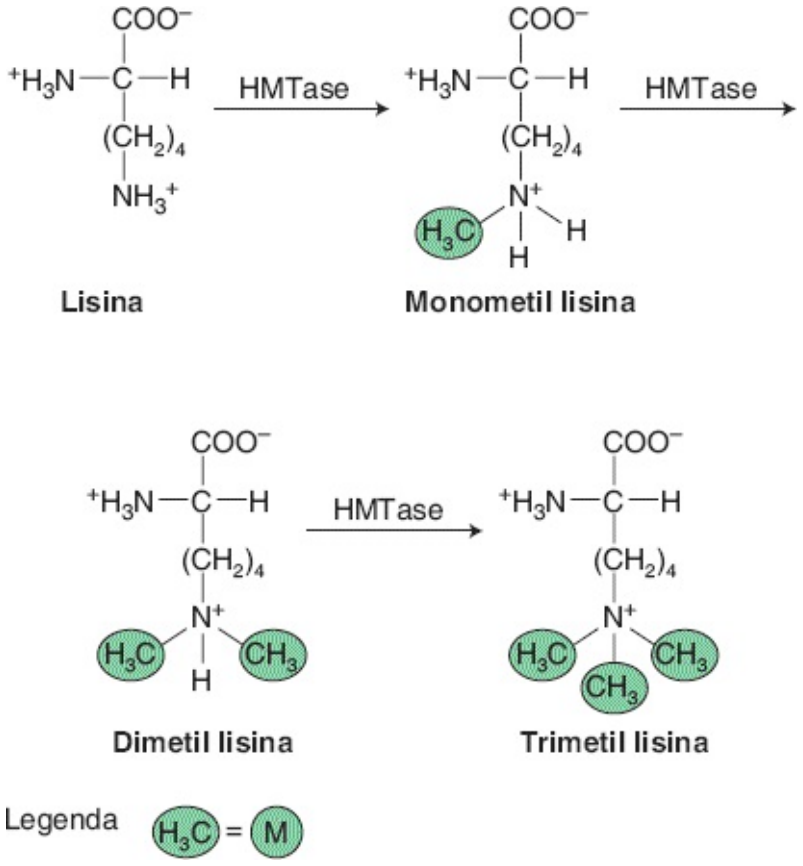


FIGURA 12.15 O recrutamento de um complexo repressor leva à repressão da transcrição. Na presença de glicose, a transcrição de *GAL1* é reprimida pela proteína Mig1, independentemente da presença de Gal4 em UAS. Mig1 se liga a um sítio entre UAS e o promotor do gene *GAL1* e recruta o complexo repressor Tup1, que recruta uma histona desacetilase, desligando a transcrição do gene.

Metilação das histonas pode ativar ou reprimir a expressão gênica

A metilação é outra modificação pós-tradução de resíduos de arginina e lisina nas caudas de histonas que pode resultar em alteração da cromatina e da expressão gênica. A enzima histona metiltransferase (HMTase) adiciona um, dois ou três grupos metil a um aminoácido específico na cauda da histona H3. Uma das reações catalisadas pela HMTase está demonstrada aqui:



O aminoácido lisina está abreviado com um “K”. Como tal, essas modificações pós-tradução da lisina 9 da histona H3 são denominadas H3K9me1, H3K9me2 e H3K9me3, respectivamente.

Contrariamente à acetilação, a adição de grupos metil pode ativar ou reprimir a expressão gênica. Relembre que a acetilação da lisina atua para neutralizar a carga positiva da histona, e, desse modo direto, ativa a expressão gênica por meio da redução das interações dos nucleossomos com o DNA, abrindo a cromatina. Contrariamente, a metilação de resíduos de lisina específicos, que não afeta a

carga, cria sítios de ligação para outras proteínas que ativam ou reprimem a expressão gênica, dependendo dos resíduos modificados. Por exemplo, a metilação do resíduo 4 da lisina da H3 [H3K4(me)] está associada à ativação da expressão gênica e está enriquecida nos nucleossomos próximos do início da transcrição. Existe um desfecho muito diferente quando H3K9 ou H3K27 são metilados. Essas modificações, que estão associadas à repressão gênica e à cromatina firmemente compactada, serão discutidas em mais detalhes posteriormente neste capítulo.

> CONCEITO-CHAVE A modificação pós-tradução das histonas está associada à ativação e à repressão da expressão gênica. Enquanto a acetilação das histonas atua diretamente para reduzir a densidade da cromatina e ativar a expressão gênica, a metilação de aminoácidos específicos das histonas cria sítios de ligação para proteínas que ativam ou reprimem a expressão gênica.

Herança das modificações de histonas e estrutura da cromatina

Uma característica importante da estrutura da cromatina é que ela pode ser herdada. Essa forma de herança recebe uma denominação – **herança epigenética** – e é definida operacionalmente como a herança dos estados da cromatina de uma geração celular para a próxima. O que essa herança significa é que, na replicação do DNA, tanto a sequência do DNA quanto a estrutura da cromatina são fielmente transmitidas para a próxima geração celular. Entretanto, contrariamente à sequência de DNA, a estrutura da cromatina pode ser alterada no curso do ciclo celular e durante gerações sucessivas de divisão celular.

Relembre que a replicação procariótica é orquestrada na forquilha de replicação pelo replissomo, uma máquina molecular que inclui duas holoenzimas DNA pol III e proteínas acessórias (ver Figura 7.20). Em eucariotos, a replicação da cromatina significa que o replissomo não apenas deve copiar a sequência de nucleotídios dos filamentos parentais, como também deve desmontar os

nucleossomos nos filamentos parentais e remontá-los nas moléculas-filhas. Durante esse processo, as histonas antigas dos nucleossomos existentes são distribuídas aleatoriamente para as moléculas-filhas e novas histonas são liberadas para o replissomo. As histonas antigas distribuídas aleatoriamente atuam como moldes para orientar a modificação das novas histonas. Desse modo, as histonas antigas, com suas caudas modificadas e as histonas novas, com caudas não modificadas, são montadas nos nucleossomos que se tornam associados a ambos os filamentos-filhos (Figura 12.16). As modificações carregadas pelas histonas antigas são responsáveis em parte pela herança epigenética. Como tal, essas antigas modificações são denominadas **marcas epigenéticas**, tendo em vista que orientam a modificação das novas histonas.

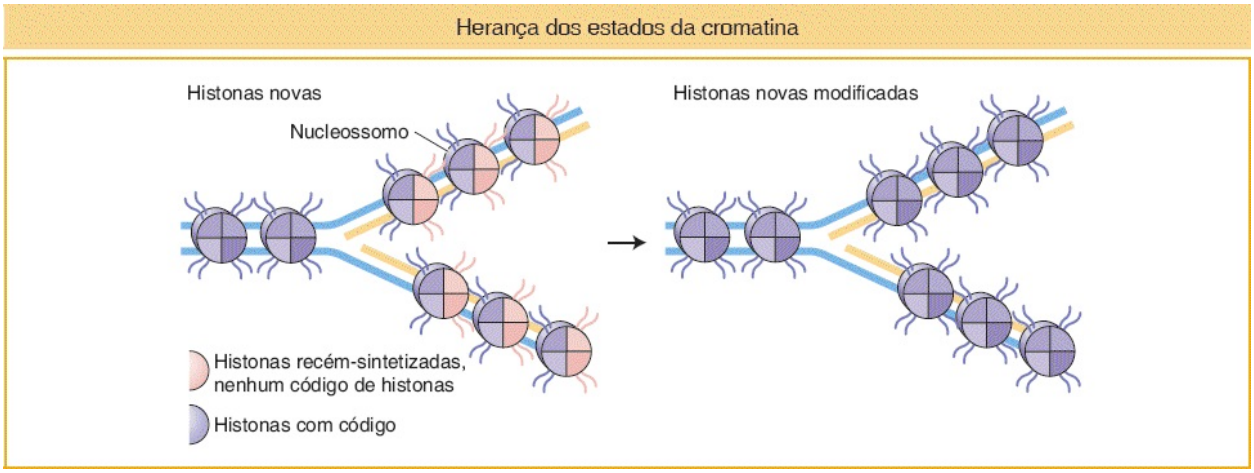


FIGURA 12.16 Na replicação, as histonas antigas (*roxo*), com seus códigos de histonas, são distribuídas aleatoriamente para os filamentos-filhos, onde direcionam a codificação das histonas adjacentes recém-montadas (*laranja*) para formar nucleossomos completos.

> CONCEITO-CHAVE O replissomo eucariótico realiza todas as funções do replissomo procariótico; além disso, ele deve desmontar e remontar os complexos de proteína-DNA denominados nucleossomos.

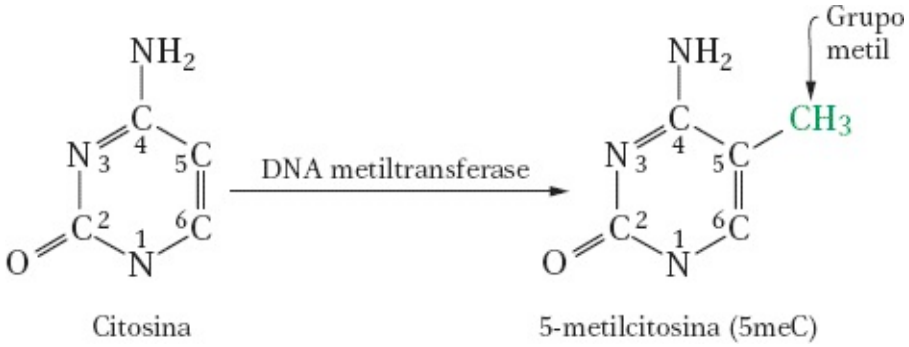
Variantes de histonas

Contrariamente às histonas comuns (também denominadas de consenso) que são

adicionadas durante a replicação do DNA, os eucariotos também apresentam outras histonas, denominadas variantes de histonas, que podem substituir as histonas de consenso que já foram montadas nos nucleossomos. Por exemplo, duas variantes em relação à histona H2 são denominadas H2A.Z e H2A.X, e uma variante da H3 é denominada CENP-A. Tendo em vista que as histonas podem ser modificadas de tantos modos, por que pode ser necessário substituir uma histona por uma variante? Embora os cientistas estejam apenas começando a compreender os diferentes papéis das variantes de histonas, um tema comum é que elas proporcionam um modo rápido para alterar a cromatina por meio da substituição de um código de histonas por outro. A CENP-A, por exemplo, substitui a H3 no DNA centromérico e acredita-se que a sua presença defina a função do centrômero. O papel da variante de histona H2A.Z na identificação do DNA danificado para o reparo rápido é discutido em detalhes no Capítulo 16.

Metilação do DNA | Outra marca hereditária que influencia a estrutura da cromatina

Existe outra marca epigenética importante na maior parte dos (mas não em todos os) eucariotos. Essa marca não é uma modificação de histonas; em vez disso, é a adição de grupos metil aos resíduos de DNA após a replicação. Uma enzima normalmente liga esses grupos metil ao carbono 5 de um resíduo de citosina específico.



Em mamíferos, o grupo metil normalmente é adicionado à citosina em um dinucleotídio CG. O padrão da metilação é denominado metilação simétrica, tendo em vista que os grupos metil estão presentes em ambos os filamentos no

mesmo contexto:

C*G
G C*

Um número extraordinário de resíduos C é metilado em mamíferos: 70 a 80% de todos os dinucleotídios CG são metilados em todo o genoma. Curiosamente, a maior parte dos dinucleotídios CG não metilados é observada em aglomerados próximos aos promotores gênicos. Essas regiões são denominadas **ilhotas de CpG**, em que o "p" representa a ligação fosfodiéster. Portanto, a metilação de C está associada a regiões inativas do genoma.

Assim como as modificações das histonas, as marcas da metilação do DNA podem ser herdadas de modo estável de uma geração celular para a próxima. A herança da metilação do DNA é mais bem-compreendida do que a herança das modificações de histonas. A replicação semiconservativa gera hélices-filhas que são metiladas apenas no filamento parental. As moléculas de DNA metiladas em apenas um filamento são denominadas **hemimetiladas**. Os grupos metil são adicionados aos filamentos não metilados por meio de DNA metiltransferases que apresentam uma alta afinidade por esses substratos hemimetilados. Essas enzimas são orientadas pelo padrão de metilação no filamento parental (Figura 12.17). Conforme você verá posteriormente no capítulo, em virtude de a metilação do DNA ser mais estável do que as modificações de histonas, com frequência ela está associada a regiões do genoma que são mantidas em um estado inativo durante todo o período de vida de um organismo. As referidas regiões serão discutidas posteriormente neste capítulo.

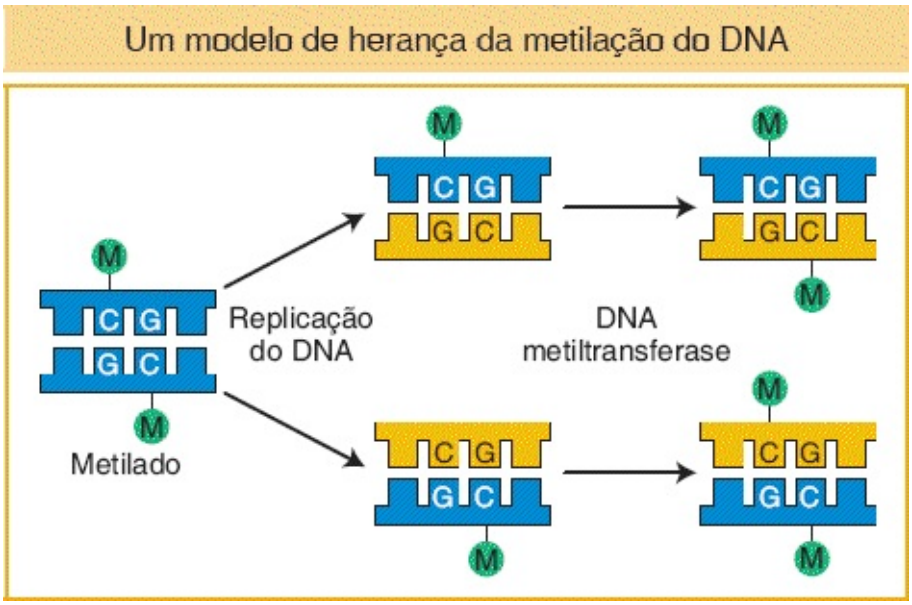


FIGURA 12.17 Após a replicação, os resíduos de dinucleotídios CG hemimetilados são totalmente metilados. Os filamentos parentais são *azuis* e o filamento-filho é *amarelo*. A letra "M" representa o grupo metil no nucleotídio C.

CONCEITO-CHAVE A estrutura da cromatina é herdada de uma geração celular para outra geração celular, tendo em vista que existem mecanismos para replicar as marcas epigenéticas associadas juntamente com o DNA. Desse modo, a informação inerente nas modificações de histonas e os padrões de metilação do DNA existentes atuam para reconstituir a estrutura da cromatina local que existia antes da síntese do DNA e da mitose. Contrariamente, as variantes das histonas podem ser utilizadas para alterar rapidamente a cromatina em uma via independente da replicação.

12.4 Ativação de genes em um ambiente de cromatina

Conforme você viu neste capítulo, a transcrição de genes eucarióticos deve ser ligada e desligada durante o período de vida de um organismo. Para compreender como os eucariotos regulam os genes durante o seu período de vida, é necessário verificar como a cromatina é alterada durante a ativação da transcrição. Além disso, o desenvolvimento de um organismo complexo requer que os níveis de

transcrição sejam regulados ao longo de uma ampla variação de atividades. Pense a respeito de um mecanismo de regulação mais como um reostato que controla o som de um iPod do que como um interruptor: em vez de um gene que produz muitas proteínas ou nenhuma proteína, ele pode produzir uma quantidade intermediária, dependendo do nível de transcrição. Em eucariotos, os níveis de transcrição se tornam finamente ajustáveis em um ambiente de cromatina por meio do agrupamento dos sítios de ligação em acentuadores. Diversos fatores de transcrição diferentes ou diversas moléculas do mesmo fator de transcrição podem se ligar a sítios adjacentes. A ligação desses fatores a sítios que estão separados na distância correta leva a um efeito amplificado ou superaditivo, sobre a ativação da transcrição. Quando um efeito é superior ao aditivo, diz-se que é **sinérgico**.

A ligação de múltiplas proteínas reguladoras aos múltiplos sítios de ligação em um acentuador pode catalisar a formação de um **acentuassomo** (*enhanceosome*), um grande complexo proteico que atua sinergicamente para ativar a transcrição. Na Figura 12.18 você pode observar como as proteínas arquitetônicas dobram o DNA para promover interações cooperativas das outras proteínas de ligação ao DNA. Nesse modo de ação do acentuassomo, a transcrição é ativada até níveis muito altos apenas quando todas as proteínas estão presentes e mantendo contato do modo correto.

Para compreender melhor o que um acentuassomo é e como ele atua sinergicamente, vejamos um exemplo específico.

Acentuassomo de β-interferona

O gene humano de β-interferona, que codifica a proteína antiviral interferona, é um dos genes mais bem-caracterizados em eucariotos. Ele normalmente está desligado, mas é ativado até níveis de transcrição muito altos na infecção viral. A chave para a ativação desse gene é a montagem de fatores de transcrição em um acentuassomo a aproximadamente 100 pb *upstream* do TATA boxe e do sítio de início da transcrição. Todas as proteínas reguladoras do acentuassomo de β-interferona se ligam à mesma face da dupla-hélice de DNA. Ligadas ao outro lado da hélice estão diversas proteínas estruturais que dobram o DNA e possibilitam

que as diferentes proteínas reguladoras se toquem e formem um complexo ativado. Quando todas as proteínas reguladoras estão ligadas e interagindo corretamente, elas formam uma "plataforma", um sítio de ligação de alta afinidade para a proteína CBP, uma proteína coativadora que também recruta o maquinário de transcrição. A grande proteína CBP também contém uma atividade de histona acetilase intrínseca que modifica os nucleossomos e facilita os níveis altos de transcrição.

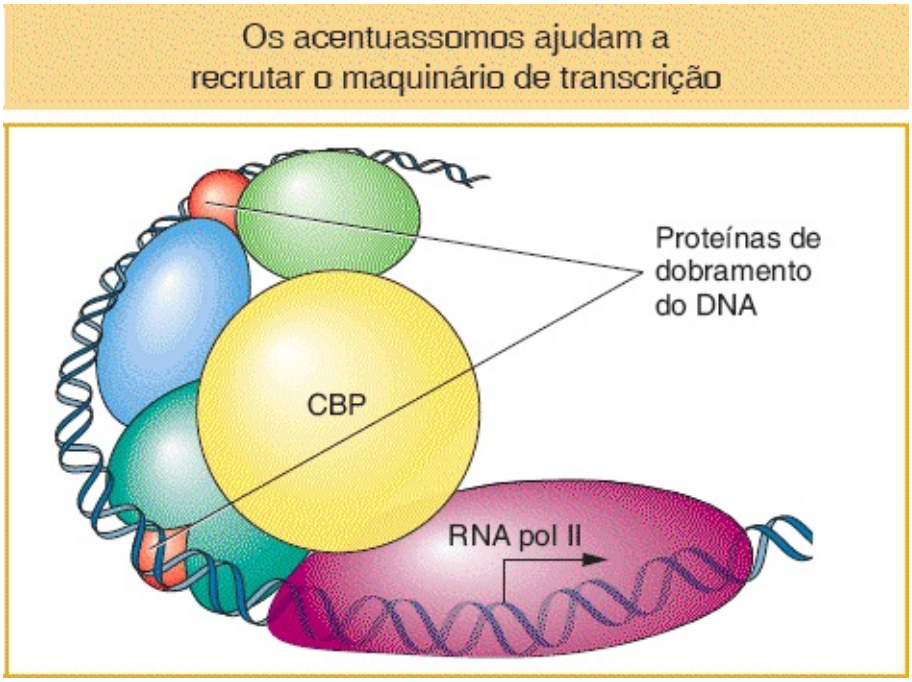


FIGURA 12.18 O acentuassomo de β-interferona. Neste caso, os fatores de transcrição recrutam um coativador (CBP), que se liga aos fatores de transcrição e à RNA polimerase II, iniciando a transcrição.

Embora o promotor de β-interferona esteja demonstrado sem nucleossomos na Figura 12.18, o acentuassomo de fato é circundado por dois nucleossomos, denominados nuc 1 e nuc 2 na Figura 12.19. Um deles, nuc 2, está estrategicamente posicionado no TATA boxe e no sítio de início da transcrição. GCN5, outro coativador, se liga e acetila os dois nucleossomos. Após a acetilação, os fatores ativadores de transcrição recrutam o coativador CBP, a holoenzima RNA pol II e o complexo de remodelagem da cromatina SWI-SNF. SWI-SNF em seguida é posicionado para afastar o nucleossomo 37 pb do TATA

boxe, tornando o TATA boxe acessível à proteína de ligação TATA e possibilitando que a transcrição seja iniciada.

As interações cooperativas ajudam a explicar diversas observações que causam perplexidade a respeito dos acentuadores. Por exemplo, elas explicam por que a mutação de qualquer fator de transcrição ou sítio de ligação reduzir dramaticamente a atividade do acentuador. Elas também explicam por que a distância entre os sítios de ligação dentro do acentuador é uma característica crítica. Além disso, os acentuadores não precisam estar próximos do sítio de início da transcrição, como é o exemplo demonstrado na Figura 12.19. Uma característica dos acentuadores é que eles podem ativar a transcrição quando estão localizados a grandes distâncias do promotor (> 50 kb), seja *upstream* ou *downstream* de um gene ou até em um íntron.

Isoladores de bloqueio de acentuador

Um elemento regulador, tal como um acentuador, que consegue atuar sobre dezenas de milhares de pares de bases pode interferir com a regulação de genes próximos. Para evitar a referida ativação promíscua, foram desenvolvidos elementos reguladores denominados **isoladores de bloqueio de acentuador**. Quando posicionados entre um acentuador e um promotor, os isoladores de bloqueio de acentuador evitam que o acentuador ative a transcrição naquele promotor. Os referidos isoladores não apresentam efeito na ativação de outros promotores que não estão separados de seus acentuadores pelo isolador (Figura 12.20). Foram propostos diversos modelos para explicar como um isolador poderia bloquear a atividade do acentuador apenas quando posicionado entre um acentuador e um promotor. Muitos dos modelos, como aquele demonstrado na Figura 12.21, propõem que o DNA está organizado em alças que contêm genes ativos. De acordo com esse modelo, os isoladores movem um promotor para uma nova alça, onde ele está protegido do acentuador.

Conforme você verá posteriormente no capítulo, os isoladores de bloqueio de acentuador são um componente fundamental de um fenômeno denominado *imprinting* genômico.

Os acentuassomos recrutam remodeladores da cromatina

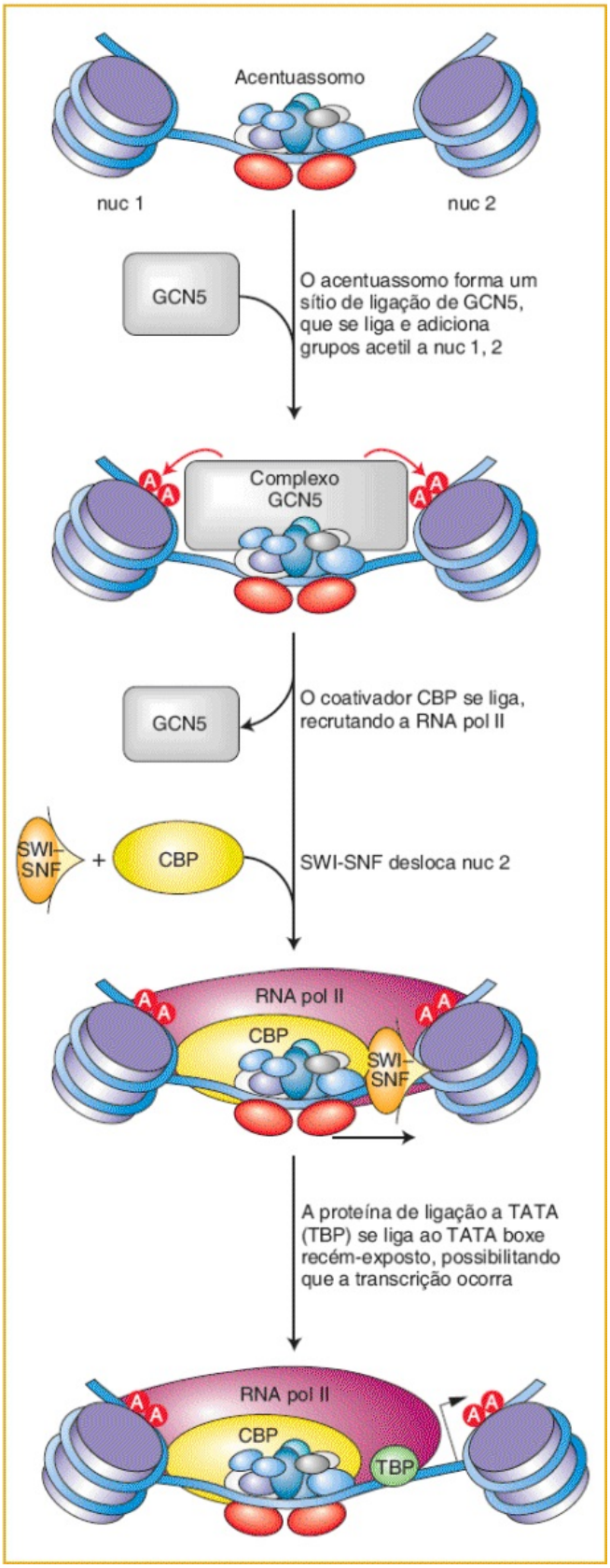


FIGURA 12.19 O acentuassomo de β-interferona atua para mover os nucleossomos por meio do recrutamento do complexo SWI-SNF.

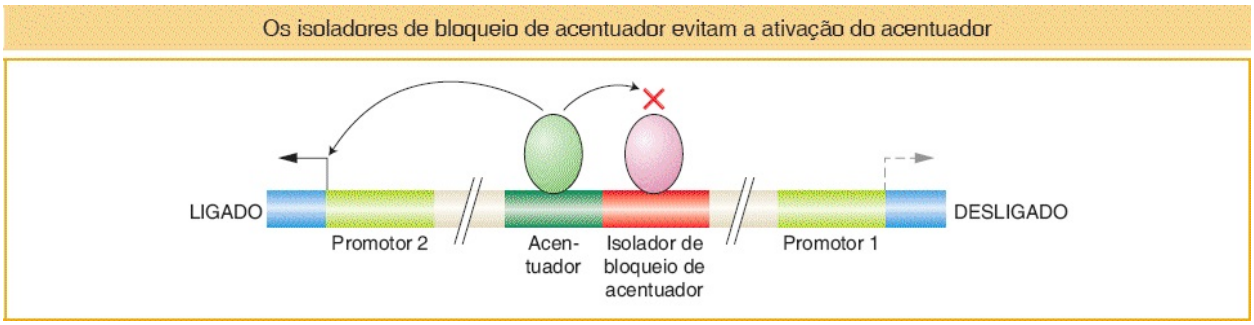


FIGURA 12.20 Os isoladores de bloqueio de acentuador evitam a ativação gênica quando posicionados entre um acentuador e um promotor.

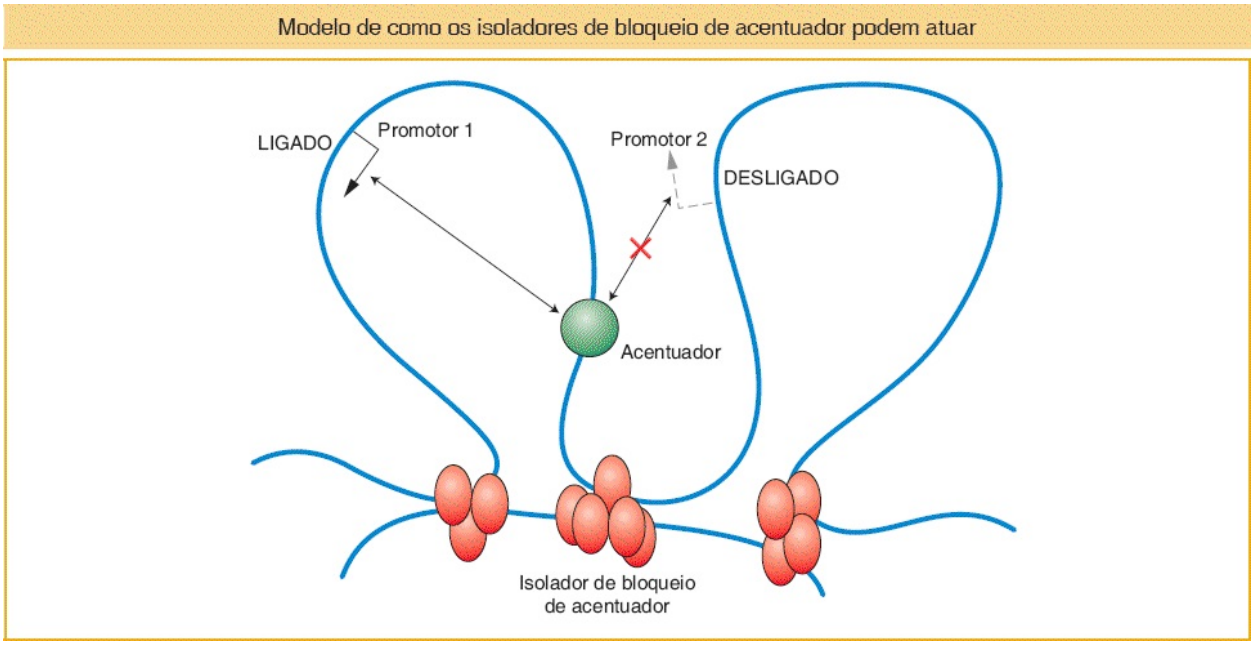


FIGURA 12.21 Uma proposta é que os isoladores de bloqueio de acentuador criam novas alças, que separam fisicamente um promotor do seu acentuador.

CONCEITO-CHAVE Os acentuadores eucarióticos podem atuar a grandes distâncias para modular a atividade do aparato de transcrição. Os acentuadores contêm sítios de ligação para muitos fatores de transcrição, que se ligam e interagem cooperativamente. Essas interações resultam em uma

diversidade de respostas, incluindo o recrutamento de coativadores adicionais e a remodelagem da cromatina.

12.5 Inativação a longo prazo de genes em um ambiente de cromatina

Até agora, vimos como os genes são ativados em um ambiente de cromatina. Entretanto, conforme declarado no início deste capítulo, a maior parte dos genes em genomas eucarióticos está desligada em qualquer momento. Um dos achados mais surpreendentes da era da genômica é que muitos genes eucarióticos são inativos durante a vida do organismo. Isso leva a duas questões que serão abordadas nesta seção. Primeiramente, por que os organismos apresentam genes que estão sempre inativos? Em segundo lugar, como os organismos mantêm os genes em um estado inativo durante todo o seu período de vida?

Um dos modelos mais úteis para a compreensão dos mecanismos que mantêm a inatividade dos genes a longo prazo diz respeito ao controle da alteração do tipo reprodutivo em leveduras. Os componentes do *locus* do tipo reprodutivo de leveduras foram introduzidos ao fim da Seção 12.2. Aqui, continuamos com a história ao enfocar no mecanismo de alteração do tipo reprodutivo, que requer que cada célula de levedura mantenha cópias inativas dos genes *a* e α em outros locais de seu genoma.

Alteração do tipo reprodutivo e silenciamento gênico

As células haploides de leveduras são capazes de alterar seu tipo reprodutivo, por vezes tão frequentemente quanto a cada ciclo celular. Desse modo, uma célula haploide de levedura de um tipo reprodutivo (digamos, **a**) formará uma colônia de ambas as células **a** e α, que podem cruzar para formar células diploides (**a**/α). Durante épocas de crise, tais como os períodos de escassez de nutrientes, cada célula diploide pode sofrer meiose e produzir quatro esporos haploides. Esse processo é vantajoso para a sobrevivência da espécie, tendo em vista que os esporos podem sobreviver em condições ambientais adversas melhor do que as

células haploides.

As análises genéticas de determinados mutantes que não conseguiam alterar seu tipo reprodutivo ou que não conseguiam reproduzir-se (eram estéreis) foram fontes de percepções-chave sobre a alteração do tipo reprodutivo. Entre os mutantes de alteração estavam diversos *loci* mutantes, incluindo o gene *HO* e os genes *HMRa* e *HMLα*. Estudos adicionais revelaram que o gene *HO* codifica uma endonuclease, enzima que cliva o DNA (ver Capítulo 10), necessária para o início da alteração. Também se observou que os *loci HMRa* e *HMLα*, que estão no mesmo cromossomo do *locus MAT*, contêm "cassetes" dos alelos *MATa* e *MATα*, respectivamente, que não são expressos. Os *loci HMR* e *HML*, portanto, são "cassetes" silenciosos. Lembre, do Capítulo 8, que um tipo de silenciamento gênico ocorre quando o dsRNA tem por alvo o complexo RISC para destruir o RNA complementar. Esse é um exemplo de **silenciamento gênico pós-transcricional**. Contrariamente, *HMRa* e *HMLα* não podem ser transcritos e, como tal, são exemplos de **silenciamento gênico transcricional**.

Duas características da alteração do tipo reprodutivo eram de interesse para os geneticistas: como as células alteram o seu tipo reprodutivo e por que *HMRa* e *HMLα* são silenciosos em relação à transcrição? A chave para a alteração é a endonuclease HO, que inicia a alteração do tipo reprodutivo ao gerar uma quebra no filamento duplo no *locus MAT*. A interconversão do tipo reprodutivo ocorre então por meio de um tipo de recombinação entre o segmento de DNA (um cassete) de um dos dois *loci* não expressos e o *locus MAT*. O resultado é a substituição do cassete antigo no *locus MAT* por um novo cassete de *HMRa* ou *HMLα*. O tipo reprodutivo resultante é o tipo MATa ou MATα, dependendo de qual gene se encontra no *locus MAT* (Figura 12.22). O cassete inserido de fato é copiado a partir do *locus HMRa* ou *HMLα*. Desse modo, a alteração é reversível, tendo em vista que a informação em relação aos cassetes *a* e α está sempre presente nos *loci HMRa* e *HMLα* e nunca é perdida. Portanto, a alteração do tipo reprodutivo fornece um exemplo de genes que precisam ser silenciados durante todo o período de vida de um organismo. Conforme você verá posteriormente neste capítulo, o silenciamento gênico durante todo o período de vida também ocorre em humanos e em todos os outros mamíferos.

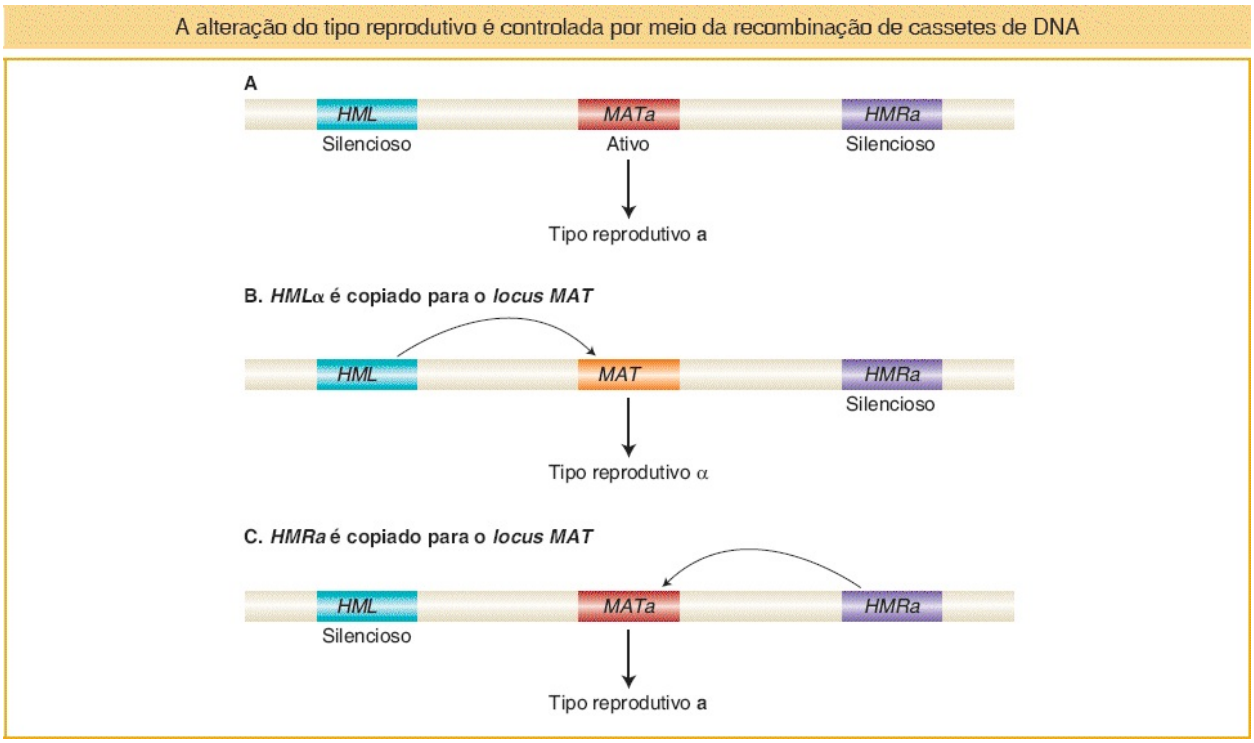


FIGURA 12.22 O cromossomo III de *S. cerevisiae* codifica três *loci* de tipos reprodutivos, mas apenas os genes no *locus MAT* são expressos. *HML* codifica um cassete silencioso de genes α e *HMR* codifica um cassete silencioso de genes *a*. A cópia de um cassete silencioso e a inserção por meio de recombinação no *locus MAT* alteram o tipo reprodutivo.

A segunda característica de alteração do tipo reprodutivo de interesse para os geneticistas é o mecanismo subjacente de silenciamento gênico. Por que os genes nos cassetes *HMRa* e *HMLα* não são expressos? Normalmente, esses cassetes são "silenciosos". Entretanto, em mutantes *SIR* (*r*eguladores de *i*nformação silenciosos), o silenciamento está comprometido, de modo que ambas as informações *a* e α são expressas. Os mutantes resultantes são estéreis. Isso significa que em leveduras normais, não mutantes, os genes nos cassetes *HMRa* e *HMLα* são capazes de serem expressos, mas não o são em virtude da ação das proteínas Sir. As proteínas Sir2, Sir3 e Sir4 formam um complexo que desempenha um papel-chave no silenciamento gênico. Sir2 é uma histona desacetilase que facilita a condensação da cromatina e ajuda a manter *HMRa* e *HMLα* em domínios da cromatina nos quais a transcrição não pode ser iniciada.

O silenciamento gênico é um processo muito diferente da repressão gênica: o

silenciamento é um efeito de posição, que depende da vizinhança na qual a informação genética está localizada. Por exemplo, um gene normalmente ativo inserido nos *loci HMR* ou *HML* será silenciado. Você aprenderá mais a respeito dos efeitos de posição posteriormente, na seção sobre a variegação por efeito de posição na mosca-das-frutas *Drosophila melanogaster*.

Comparação da heterocromatina e da eucromatina

Examinaremos por que o silenciamento gênico a longo prazo, do tipo que silencia *HMLα* e *HMRa*, é um processo diferente da repressão gênica. Para tanto, é importante observar que a cromatina não é uniforme em todos os cromossomos: determinadas regiões cromossômicas estão agrupadas em uma cromatina altamente condensada, denominada **heterocromatina**. Outros domínios estão acondicionados em uma cromatina menos condensada, denominada **eucromatina** (ver Figura 12.11 B). A condensação da cromatina é alterada durante o período do ciclo celular. A cromatina das células que entram em mitose se torna altamente condensada na medida em que os cromossomos se alinham no preparo para a divisão celular. Após a divisão celular, as regiões que formam a heterocromatina permanecem condensadas, especialmente ao redor dos centrômeros e dos telômeros (denominada **heterocromatina constitutiva**), enquanto as regiões que formam a eucromatina se tornam menos condensadas. Conforme observado em relação ao exemplo de β-interferona (ver Figura 12.19), a cromatina dos genes ativos pode ser alterada em resposta ao estágio do desenvolvimento ou às condições ambientais.

A principal distinção entre a heterocromatina e a eucromatina é que a primeira contém poucos genes, enquanto a última é rica em genes. Mas o que é a heterocromatina, se não genes? A maior parte do genoma eucariótico é composta por sequências repetitivas que não codificam proteína ou RNA estrutural (ver Capítulo 14). Portanto, diz-se que os nucleossomos densamente compactados da heterocromatina (organizados em fibras de cromatina de 30 nm; ver Figura 12.11 A) formam uma estrutura “fechada”, que é amplamente inacessível às proteínas reguladoras e inóspita para a atividade gênica. Contrariamente, a eucromatina, com seus nucleossomos mais amplamente espaçados (organizados em fibras de 10

nm; ver Figura 12.11 A), adota uma estrutura “aberta” que possibilita a transcrição.

> CONCEITO-CHAVE A cromatina dos eucariotos não é uniforme. As regiões heterocromáticas altamente condensadas apresentam menos genes e frequências de recombinação mais baixas do que as regiões eucromáticas menos condensadas.

Variegação por efeito de posição em Drosophila revela vizinhanças genômicas

Muito antes de os *loci* de reprodução silenciosos de leveduras terem sido descritos, o geneticista Hermann Muller descobriu um fenômeno genético interessante enquanto estudava a *Drosophila*: existem vizinhanças cromossômicas que podem silenciar genes que são experimentalmente “realocados” para regiões adjacentes do cromossomo. Nesses experimentos, as moscas foram irradiadas com raios X para induzir mutações em suas células germinativas. A progênie das moscas irradiadas foi triada em relação a fenótipos incomuns. Uma mutação no gene *white*, próximo à extremidade do cromossomo X, resultará em uma progênie com olhos brancos, em vez da cor vermelha do tipo selvagem. Parte da progênie apresentou olhos muito incomuns, com manchas brancas e vermelhas. O exame citológico revelou um rearranjo cromossômico nas moscas mutantes: no cromossomo X encontrava-se presente uma inversão de um pedaço do cromossomo que carrega o gene *white* (Figura 12.23). As inversões e outros rearranjos cromossômicos serão discutidos no Capítulo 17. Nesse rearranjo, o gene *white*, que normalmente está localizado em uma região eucromática do cromossomo X, agora se encontra próximo do centrômero heterocromático. Em algumas células, a heterocromatina pode se “difundir” para a eucromatina vizinha e silenciar o gene *white*. As manchas de tecido branco nos olhos são derivadas dos descendentes de uma única célula na qual o gene *white* foi silenciado e permanece silenciado durante as futuras divisões celulares. Contrariamente, as

manchas vermelhas têm origem em células nas quais a heterocromatina não se difundiu para o gene *white* e, assim, esse gene permanece ativo em todos os seus descendentes. A existência de manchas vermelhas e brancas em células do olho de um único organismo ilustra dramaticamente duas características do silenciamento epigenético. Primeiramente, que a expressão de um gene pode ser reprimida em virtude de sua posição no cromossomo, em vez de por uma mutação na sequência de seu DNA. Em segundo lugar, que o silenciamento epigenético pode ser herdado de uma geração celular para a próxima.

Os achados de estudos subsequentes em *Drosophila* e leveduras demonstraram que muitos genes ativos são silenciados nesse modo mosaico quando eles são realocados para regiões vizinhas (próximo de centrômeros ou telômeros) heterocromáticas. Portanto, a capacidade da heterocromatina de se difundir para a eucromatina e silenciar genes é uma característica comum a muitos organismos. Esse fenômeno foi denominado **variegação por efeito de posição (PEV)**. Ele fornece evidências poderosas de que a estrutura da cromatina é capaz de regular a expressão de genes – nesse caso, determinar se genes com sequências de DNA idênticas serão ativos ou silenciados.

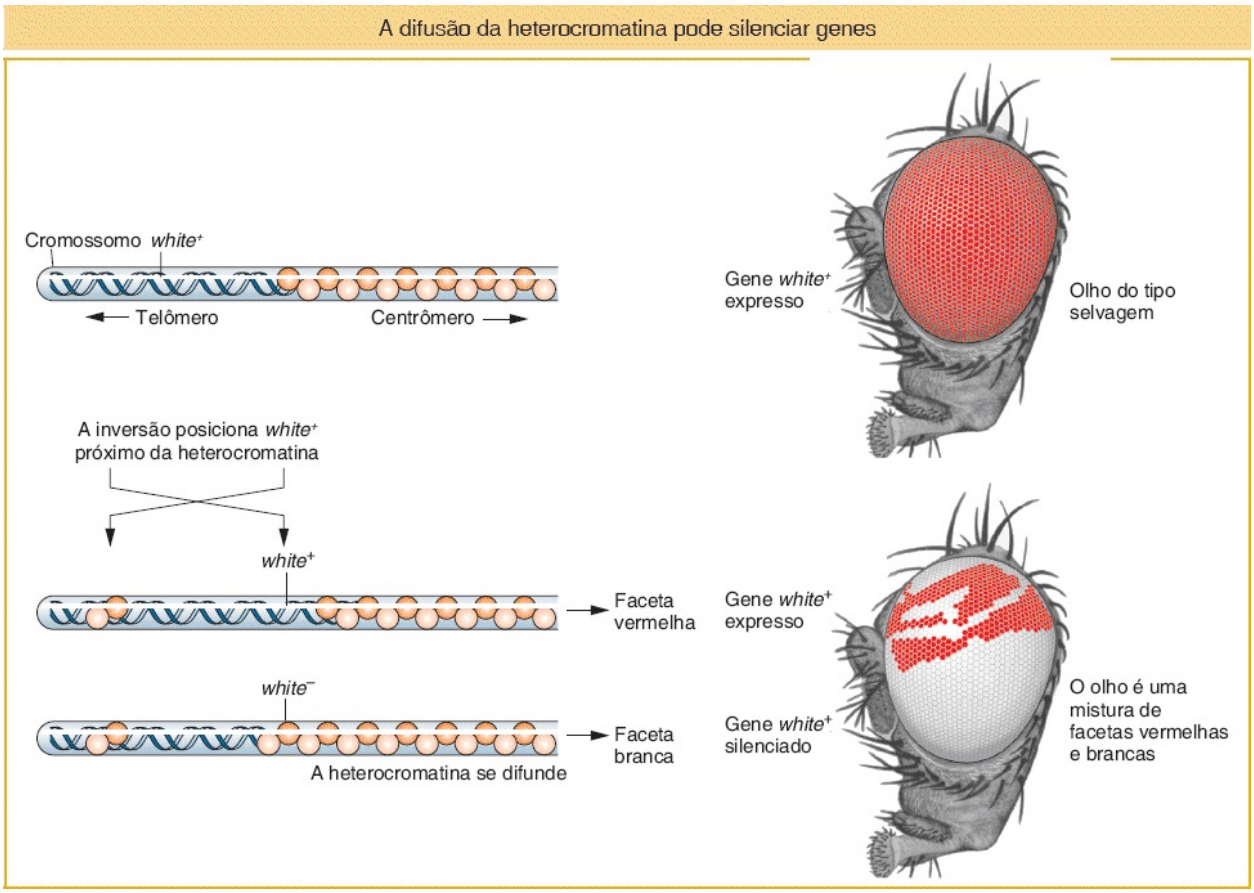


FIGURA 12.23 O rearranjo cromossômico produz variegação por efeito de posição. A inversão cromossômica posiciona o alelo *white* do tipo selvagem próximo da heterocromatina. A difusão da heterocromatina silencia o alelo. As facetas oculares são brancas, em vez do vermelho do tipo selvagem, onde o alelo foi silenciado.

> CONCEITO-CHAVE Os genes ativos que são realocados para adjacências genômicas que são heterocromáticas podem ser silenciados se a heterocromatina se difundir para os genes.

Análise genética de PEV revela proteínas necessárias para a formação de heterocromatina

O QUE OS GENETICISTAS FAZEM ATUALMENTE

Os geneticistas deduziram que a PEV poderia ser explorada para identificar as

proteínas necessárias para a formação da heterocromatina. Para essa finalidade, eles isolaram mutações em um segundo *locus* cromossômico que suprimia ou acentuava o padrão variegado (Figura 12.24). Os supressores da variegação [denominados *Su(var)*] são genes que, quando mutados, reduzem a difusão da heterocromatina, o que significa que os produtos do tipo selvagem desses genes são necessários para a difusão. De fato, os alelos *Su(var)* comprovaram ser um grande tesouro para os cientistas interessados nas proteínas que são necessárias para estabelecer e manter o estado heterocromático inativo. Entre os mais de 50 produtos dos genes de *Drosophila* identificados por essas triagens estava a proteína heterocromatina 1 (HP-1), a qual anteriormente observou-se estar associada aos telômeros e centrômeros heterocromáticos. Portanto, faz sentido que uma mutação no gene que codifica HP-1 mostre-se como um alelo *Su(var)*, tendo em vista que de algum modo a proteína é necessária para gerar a estrutura da cromatina de ordem superior associada à heterocromatina.

Mas por que a HP-1 se liga a algumas regiões do DNA e não a outras? A resposta a isso surgiu com a descoberta de que outro gene *Su(var)* codificava uma metiltransferase que adiciona grupos metil à lisina 9 na cauda da histona H3 (denominada histona H3 metiltransferase ou HMTase). Relembre que o H3K9me está associado à repressão da expressão gênica (ver Figura 12.13 B). Isso em virtude de a cromatina modificada desse modo se ligar às proteínas HP-1, que em seguida se associam para formar heterocromatina. Proteínas semelhantes à HP-1 e à HMTase foram isoladas em diversos grupos taxonômicos, sugerindo a conservação de uma importante função eucariótica.

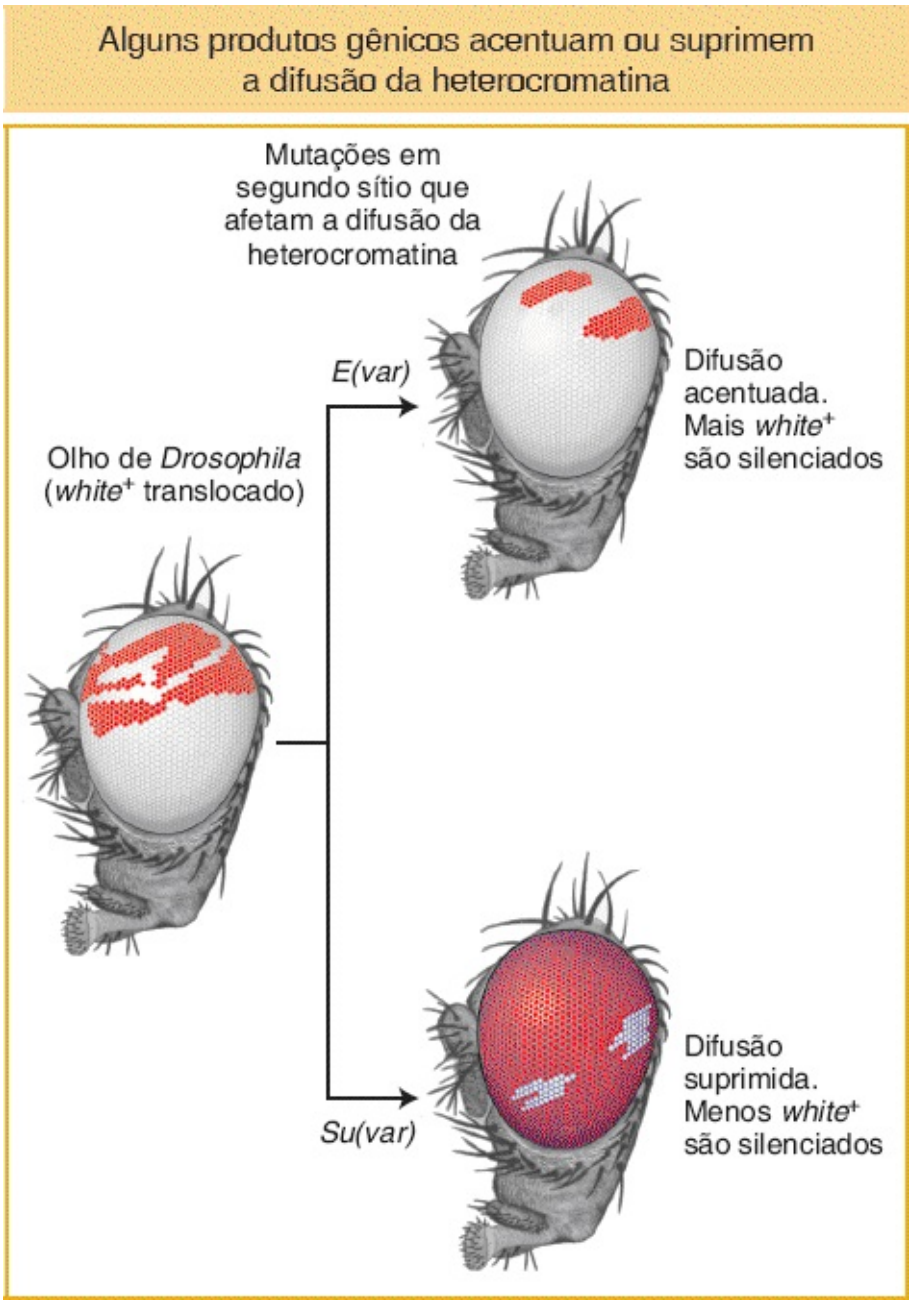


FIGURA 12.24 Foram utilizadas mutações para identificar genes que suprimem, *Su(var)*, ou acentuam, *E(var)*, a variegação por efeito de posição.

Observamos que regiões transcritas ativamente estão associadas a nucleossomos cujas caudas de histonas são hiperacetiladas e que ativadores de transcrição tais como *GCN5* codificam uma atividade de histona acetiltransferase. Conforme já discutido, as marcas de acetil também podem ser removidas das

histonas por histona desacetilases. De modo semelhante, a cromatina composta por nucleossomos que são metilados em H3K9 e ligados à proteína HP-1 contém marcas epigenéticas que estão associadas à heterocromatina. Atualmente, os cientistas são capazes de separar a heterocromatina e a eucromatina e analisar as diferenças nas modificações das histonas e nas proteínas ligadas. O procedimento utilizado, *i*muno*p*recipitação da *c*romatina (ChIP), está descrito no Capítulo 14.

A Figura 12.25 ilustra que, na ausência de quaisquer barreiras, a heterocromatina pode se espalhar para regiões adjacentes e inativar genes em algumas células, mas não em outras. Isso pode ser o que está ocorrendo com o gene *white* de *Drosophila* quando ele é translocado para próximo do domínio de heterocromatina associado às extremidades do cromossomo. Mas a difusão da heterocromatina pode ser interrompida? Pode-se imaginar que a difusão da heterocromatina para regiões gênicas ativas pode ser desastrosa para um organismo, tendo em vista que os genes ativos seriam silenciados na medida em que fossem convertidos em heterocromatina. Para evitar esse possível desastre, o genoma contém elementos de DNA denominados **isoladores de barreira**, que previnem a difusão da heterocromatina por meio da criação de um ambiente local que não é favorável à formação de heterocromatina. Por exemplo, um isolador de barreira pode se ligar a HAT e, ao fazer isso, assegurar que as histonas adjacentes sejam hiperacetiladas. Um modelo sobre como um isolador de barreira pode atuar para "proteger" uma região da eucromatina da conversão em heterocromatina está demonstrado na Figura 12.26.

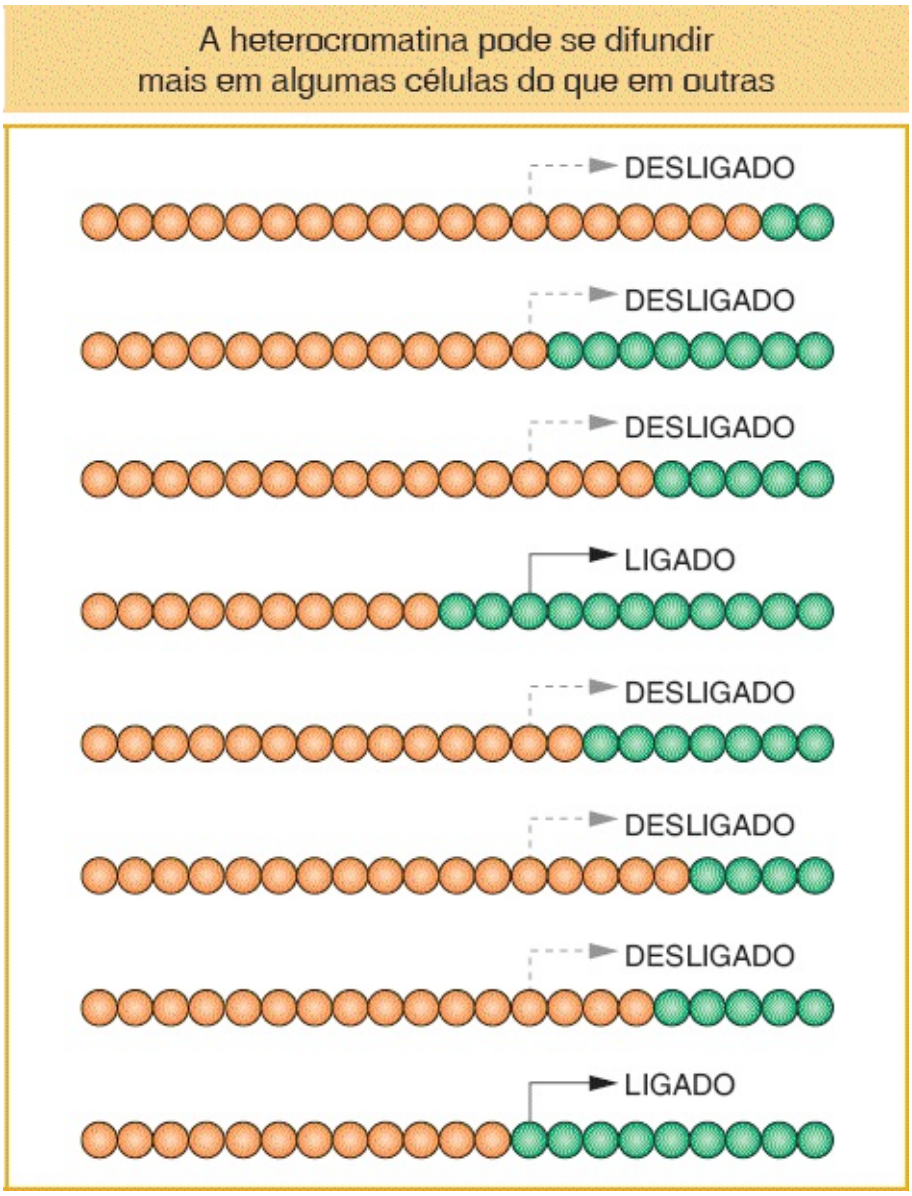


FIGURA 12.25 A difusão da heterocromatina na eucromatina adjacente é variável. Em quatro células diploides geneticamente idênticas, a heterocromatina se difunde o suficiente para o nocaute gênico em alguns cromossomos, mas não em outros. A heterocromatina e a eucromatina estão representadas pelas *esferas laranja* e *verdes*, respectivamente.

CONCEITO-CHAVE O isolamento das proteínas críticas necessárias para a formação de heterocromatina, incluindo HP-1 e HMTase, foi possibilitado pelo isolamento de linhagens mutantes de *Drosophila* que suprimiam ou acentuavam PEV.

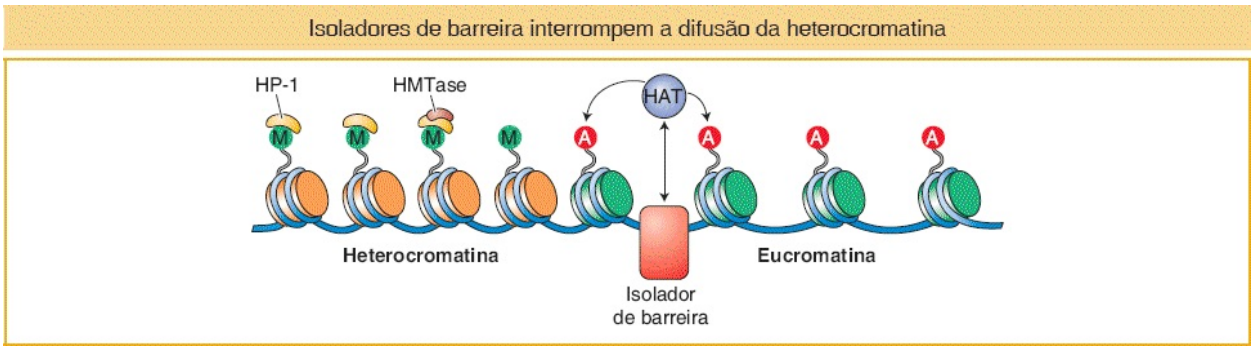


FIGURA 12.26 Neste modelo, isoladores de barreira recrutam atividades enzimáticas, como da histona acetiltransferase (HAT), que promovem a formação de eucromatina. A letra "M" faz referência à metilação e a letra "A", à acetilação.

12.6 Silenciamento gênero-específico de genes e cromossomos inteiros

Até agora, discutimos os domínios cromossômicos que estão abertos ou condensados em todos os membros de uma espécie. Nesta seção, consideramos dois fenômenos genéticos disseminados em mamíferos que dependem do sexo do indivíduo. Nesses casos, genes específicos ou até um cromossomo inteiro são silenciados durante toda a vida de um organismo. Entretanto, contrariamente aos exemplos anteriores, esses genes ou cromossomos são silenciados no sexo masculino ou feminino, mas não em ambos.

Imprinting genômico explica alguns padrões de herança incomuns

O fenômeno de ***imprinting* genômico** foi descoberto há aproximadamente 20 anos em mamíferos. No *imprinting* genômico, determinados genes autossômicos apresentam padrões de herança incomuns. Por exemplo, um alelo *Igf2* é expresso em um camundongo apenas se ele for herdado do pai do camundongo – um exemplo de ***imprinting* materno**, tendo em vista que a cópia do gene derivada da mãe é inativa. Contrariamente, um alelo *H19* de camundongo é expresso apenas se for herdado da mãe; o *H19* é um exemplo de ***imprinting* paterno**, tendo em vista que a cópia paterna é inativa. A consequência do *imprinting* parental é que os genes imprintados são expressos como se fossem a única cópia do gene presente

na célula, embora existam duas. De importância, nenhuma alteração é observada nas sequências de DNA dos genes imprintados; ou seja, o gene idêntico pode ser ativo ou inativo na progênie, dependendo de ter sido herdado da mãe ou do pai. Isso, então, representa um fenômeno epigenético.

Se a sequência do DNA do gene não está correlacionada com a atividade, o que está correlacionado? A resposta é que durante o desenvolvimento dos gametas, grupos metil são adicionados ao DNA nas regiões reguladoras de genes imprintados apenas em um sexo. Vimos anteriormente que o DNA de genes que estão desligados durante toda a vida normalmente é altamente metilado. Entretanto, é importante observar que a metilação do DNA é uma de diversas marcas epigenéticas associadas à inativação gênica a longo prazo. Outras marcas incluem a metilação de aminoácidos de histonas específicos, incluindo H3K27me1.

Voltemos novamente aos genes *Igf2* e *H19* de camundongos para verificar como o *imprinting* atua no nível molecular. Esses dois genes estão localizados em um grupo de genes imprintados no cromossomo 7 de camundongo. Estima-se que existam 100 genes imprintados em camundongo e a maior parte é observada em grupos que compreendem de 3 a 11 genes imprintados. (Os seres humanos apresentam a maior parte dos mesmos genes imprintados de camundongos.) Em todos os casos examinados, existe um padrão específico de metilação do DNA para cada cópia de um gene imprintado. Em relação ao grupo *Igf2-H19*, uma região específica do DNA localizada entre os dois genes (Figura 12.27) é metilada em células germinativas masculinas e não metilada em células germinativas femininas. Essa região é denominada região de controle de *imprinting* (ICR). Portanto, a metilação da ICR leva ao estado ativo de *Igf2* e ao estado inativo de *H19*, enquanto a ausência de metilação leva ao contrário.

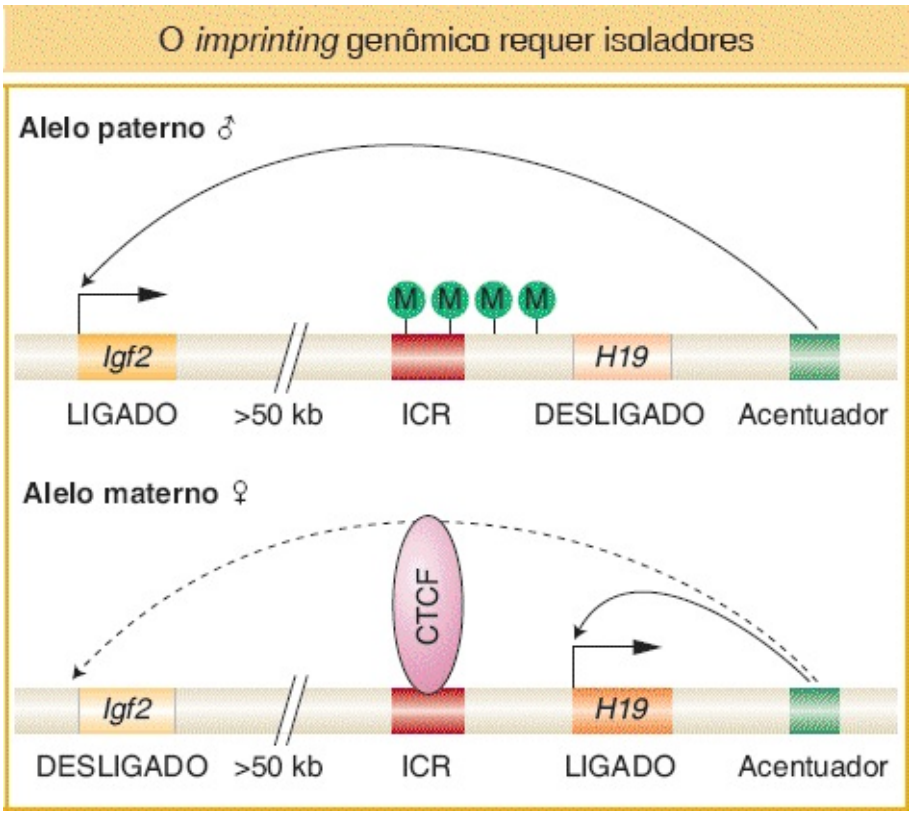


FIGURA 12.27 *Imprinting* genômico no camundongo. A região de controle do *imprinting* (ICR) não é metilada nos gametas femininos e pode se ligar a um dímero CTCF, formando um isolador que bloqueia a ativação do acentuador de *Igf2*. A metilação (M) de ICR nas células germinativas masculinas impede a ligação da CTCF, mas também impede a ligação de outras proteínas ao promotor *H19*.

Como a metilação controla qual dos dois genes está ativo? A metilação atua como um bloqueio à ligação das proteínas necessárias para a transcrição. Apenas a ICR não metilada (feminina) pode ligar-se a uma proteína reguladora denominada CTCF. Quando ligada, CTCF atua como um isolador de bloqueio de acentuador que impede a ativação do acentuador da transcrição de *Igf2*. Entretanto, o acentuador nas fêmeas ainda consegue ativar a transcrição de *H19*. Nos machos, CTCF não consegue se ligar à ICR e o acentuador consegue ativar a transcrição de *Igf2* (lembre que os acentuadores podem atuar a grandes distâncias). Entretanto, o acentuador não consegue ativar *H19*, tendo em vista que a região metilada se estende para o promotor *H19*. O promotor metilado não consegue se ligar às proteínas necessárias para a transcrição de *H19*.

Portanto, vemos como um isolador de bloqueio de acentuador (nesse caso,

CTCF ligada a parte de ICR) impede que o acentuador ative um gene distante (nesse caso, *Igf2*). Além disso, vemos que o sítio de ligação de CTCF é metilado apenas em cromossomos derivados genitor do sexo masculino. A metilação do sítio de ligação de CTCF impede a ligação de CTCF nos machos e possibilita que o acentuador ative *Igf2*.

Observe que o *imprinting* parental pode afetar amplamente a análise de heredogramas. Tendo em vista que o alelo herdado de um genitor é inativo, uma mutação no alelo herdado do outro genitor aparentará ser dominante, embora, de fato, o alelo seja expresso apenas em virtude de um dos dois homólogos ser ativo em relação a esse gene. A Figura 12.28 demonstra como uma mutação em um gene imprintado pode apresentar desfechos diferentes sobre o fenótipo do organismo se herdado do genitor do sexo masculino ou do sexo feminino.

São necessárias muitas etapas para o *imprinting* (Figura 12.29). Logo após a fertilização, os mamíferos reservam células que se tornarão suas células germinativas. Os *imprints* são apagados antes que as células germinativas sejam formadas. Sem a sua marca distinta de metilação do DNA, diz-se agora que esses genes são *epigeneticamente equivalentes*. Tendo em vista que essas células germinativas primordiais se tornam gametas totalmente formados, os genes imprintados recebem a marca específica do sexo que determinará se o gene será ativo ou silencioso após a fertilização.

Mas o que dizer sobre Dolly e outros mamíferos clonados?

Muitos acreditavam que o *imprinting* genômico levaria a uma exigência da participação tanto da célula germinativa masculina quanto da feminina no desenvolvimento embrionário de mamíferos. Ou seja, os gametas masculinos e femininos contêm diferentes subconjuntos de genes imprintados; assim, as células germinativas de ambos os sexos devem participar para que o embrião apresente um complemento total de genes imprintados ativos. Por que, então, mamíferos tais como Dolly e, mais recentemente, porcos, gatos, cães e vacas clonados, que foram derivados de núcleos somáticos, são capazes de sobreviver e desenvolver-se? Afinal, conforme já observado, mesmo a mutação de um único gene imprintado pode ser letal ou levar a uma doença grave.

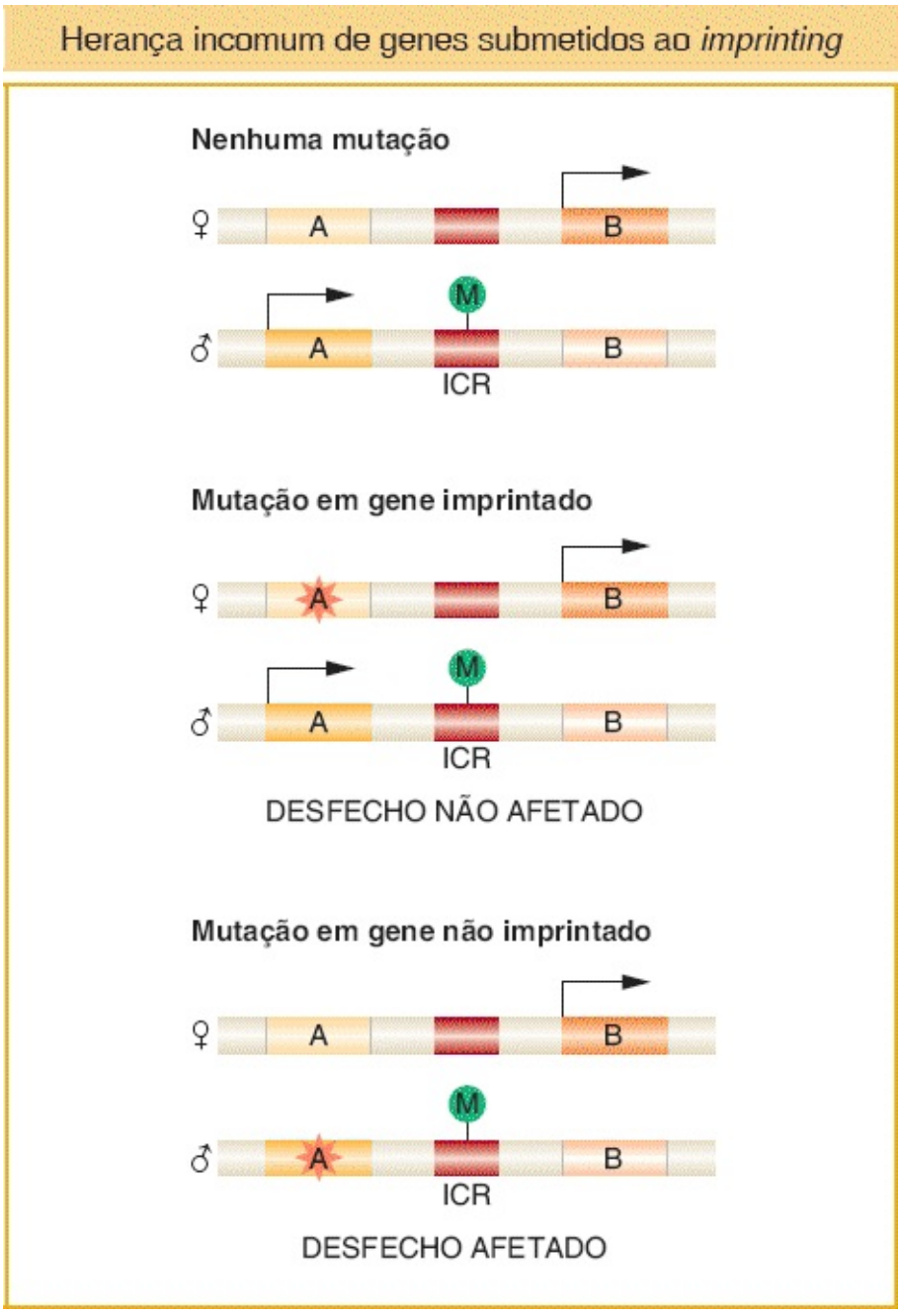


FIGURA 12.28 Uma mutação (representada por uma *estrela laranja*) no gene A não apresentará efeitos se for herdada do sexo masculino. Abreviações: M = Metilação; ICR = Região de controle do imprinting.

Nesse ponto, os cientistas não compreendem o motivo da clonagem de muitas espécies de mamíferos ter obtido sucesso. Entretanto, apesar desses sucessos, a clonagem é extremamente ineficiente em todas as espécies testadas. Em relação à maior parte dos experimentos, um clone de sucesso é um evento extremamente

raro, que requer centenas, até mesmo milhares de tentativas. Poderia ser argumentado que a falha da maior parte dos embriões clonados em se desenvolver em organismos viáveis é um atestado da importância dos mecanismos epigenéticos de regulação gênica em eucariotos. Como tal, isso ilustra como o conhecimento sobre a sequência completa do DNA de todos os genes em um organismo é apenas uma primeira etapa na compreensão sobre como os genes eucarióticos são regulados.

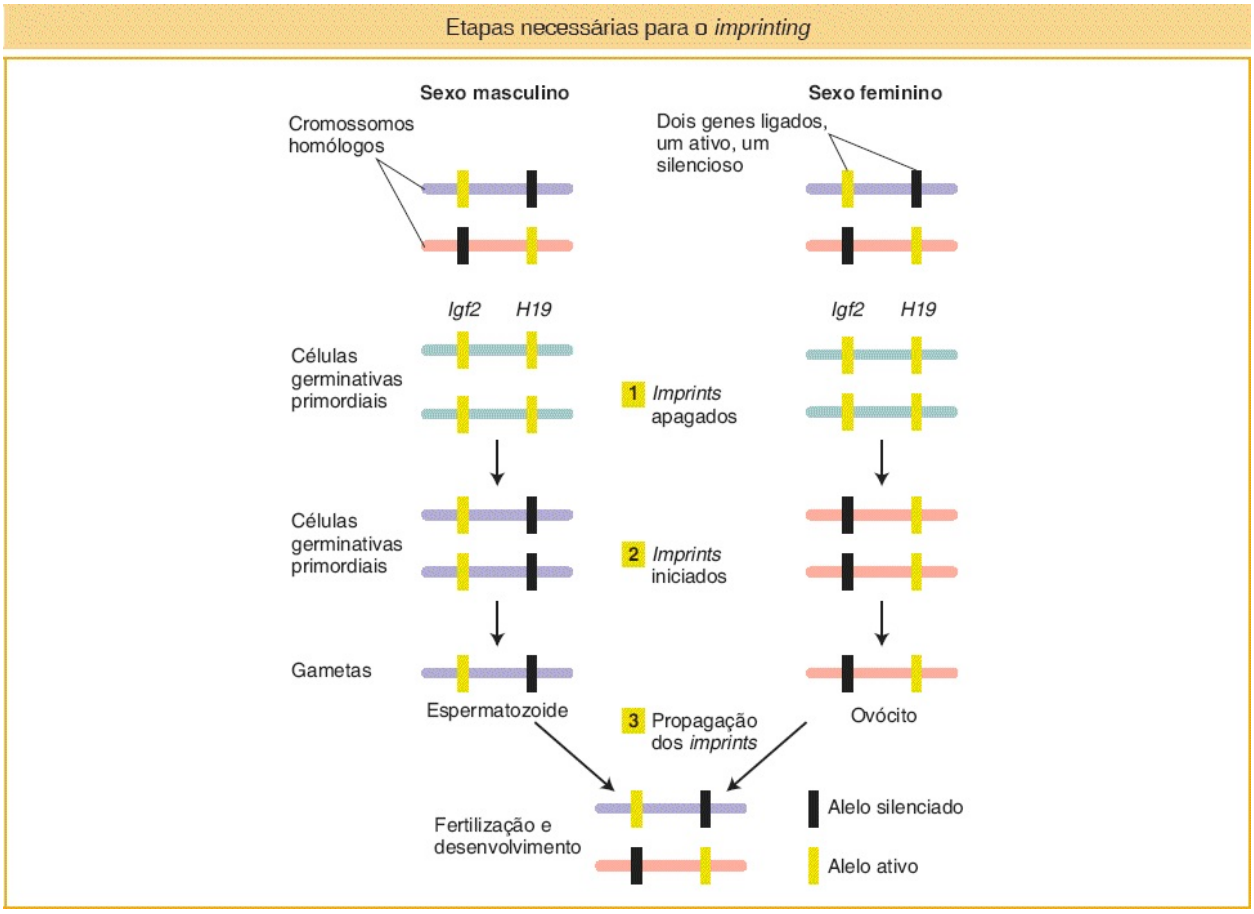


FIGURA 12.29 Como *Igf2* e *H19* são imprintados de modo diferencial nos sexos masculino e feminino.

Silenciamento de um cromossomo inteiro | Inativação do cromossomo X

O fenômeno epigenético denominado inativação do cromossomo X tem intrigado cientistas há décadas. No Capítulo 17, você aprenderá a respeito dos efeitos do

número de cópias gênicas sobre o fenótipo de um organismo. Por enquanto, você precisa saber apenas que o número de transcritos produzidos por um gene normalmente é proporcional ao número de cópias daquele gene em uma célula. Os mamíferos, por exemplo, são diploides e apresentam duas cópias de cada gene localizadas em seus autossomos. Para a vasta maioria dos genes, ambos os alelos são expressos. Portanto, todos os indivíduos estão produzindo aproximadamente o mesmo número de transcritos em relação a esses genes, proporcional a duas cópias dos genes.

Entretanto, existe uma exceção a essa generalização. Todos os indivíduos não produziriam o mesmo número de transcritos de genes localizados nos cromossomos sexuais se ambos os cromossomos X fossem expressos no sexo feminino. Conforme discutido no Capítulo 2, o número de cromossomos sexuais X e Y difere entre os sexos, com as fêmeas de mamíferos apresentando dois cromossomos X e os machos apenas um. Acredita-se que o cromossomo X de mamíferos contenha aproximadamente 1.000 genes. As fêmeas apresentam o dobro de cópias dos genes ligados ao X em relação aos machos e expressariam o dobro de transcritos desses genes do que os machos se não houvesse um mecanismo para corrigir esse desequilíbrio. (A ausência de um cromossomo Y não é um problema para os indivíduos do sexo feminino, tendo em vista que os pouquíssimos genes nesse cromossomo são necessários apenas para o desenvolvimento da masculinidade.) Dizemos que as fêmeas produzem duas doses de transcritos para cada dose produzida pelos machos.

Esse desequilíbrio de dosagem é corrigido por meio de um processo denominado **compensação de dose**, que torna a quantidade da maior parte dos produtos gênicos das duas cópias do cromossomo X no sexo feminino equivalente à dose única do cromossomo X no sexo masculino. Em mamíferos, a equivalência de doses é alcançada por meio da inativação aleatória de um dos dois cromossomos X em cada célula em um estágio inicial no desenvolvimento. Esse estado inativo em seguida é propagado para todas as células da progênie. (Na linhagem germinativa, o segundo cromossomo X se torna reativado na ovogênese.) O cromossomo inativado, denominado **corpúsculo de Barr**, pode ser observado no núcleo como uma estrutura heterocromática altamente condensada e

fortemente corada.

A inativação do cromossomo X é um exemplo de herança epigenética. Primeiramente, a maior parte dos genes no cromossomo X inativado (denominado Xi) é silenciada e o cromossomo apresenta marcas epigenéticas associadas à heterocromatina, incluindo H3K9me, hipoacetilação de histonas e hipermetilação de seu DNA. Em segundo lugar, a maior parte dos genes no cromossomo inativado, mas não todos, permanece inativa em todos os descendentes dessas células, ainda que a própria sequência do DNA esteja inalterada.

O mecanismo que converte um cromossomo X totalmente funcional em heterocromatina é o objeto de investigações atuais. O processo está bem-caracterizado em camundongos e a inativação do cromossomo X naquele organismo compartilha muitas características com a inativação do cromossomo X em células somáticas femininas humanas. Ambos apresentam um *locus* no cromossomo X denominado centro de inativação do X (abreviado *Xic*), que produz um RNA não codificador de proteína (ncRNA; ver Capítulo 8) de 17 kb denominado Xist. Acredita-se que o Xist seja transcrito a partir de apenas um cromossomo no início do desenvolvimento dos embriões de fêmeas de camundongos. O cromossomo que produz Xist se torna inativo na medida em que o Xist reveste especificamente a região central desse cromossomo, levando à formação de heterocromatina. Não se sabe como o Xist está localizado em um cromossomo, nem o modo como ele aciona a conversão para heterocromatina.

Um modelo interessante em relação ao modo como a transcrição de um ncRNA pode influenciar a estrutura da cromatina no Xi está demonstrado na Figura 12.30. De acordo com esse modelo, na medida em que um ncRNA é transcrito pela RNA pol II, proteínas se ligam especificamente às suas sequências e catalisam as modificações de histona que iniciam a formação de heterocromatina. Desse modo, os ncRNA atuam como amarras para recrutar proteínas modificadoras da cromatina para o cromossomo X a partir do qual ele é transcrito.

> CONCEITO-CHAVE Em relação à maior parte dos organismos diploides, ambos os alelos de um gene são expressos de modo independente. O

imprinting genômico e a inativação do X são exemplos de expressão monoalélica. Nesses casos, mecanismos epigenéticos silenciam um único *locus* cromossômico ou uma cópia de um cromossomo inteiro, respectivamente.

12.7 Repressão gênica pós-transcricional pelos miRNA

O Xist é um exemplo da classe em rápido crescimento de RNA funcionais (ver Capítulo 8). Os RNA funcionais não codificam proteínas; em vez disso, realizam uma diversidade de tarefas que exploram a complementaridade do RNA-RNA e do RNA-DNA. Os RNA funcionais discutidos aqui contêm sequências específicas que direcionam as proteínas ou os complexos proteicos para locais na célula onde seus serviços são necessários. Por exemplo, o Xist atua para direcionar proteínas envolvidas na formação da heterocromatina para um dos dois cromossomos X.

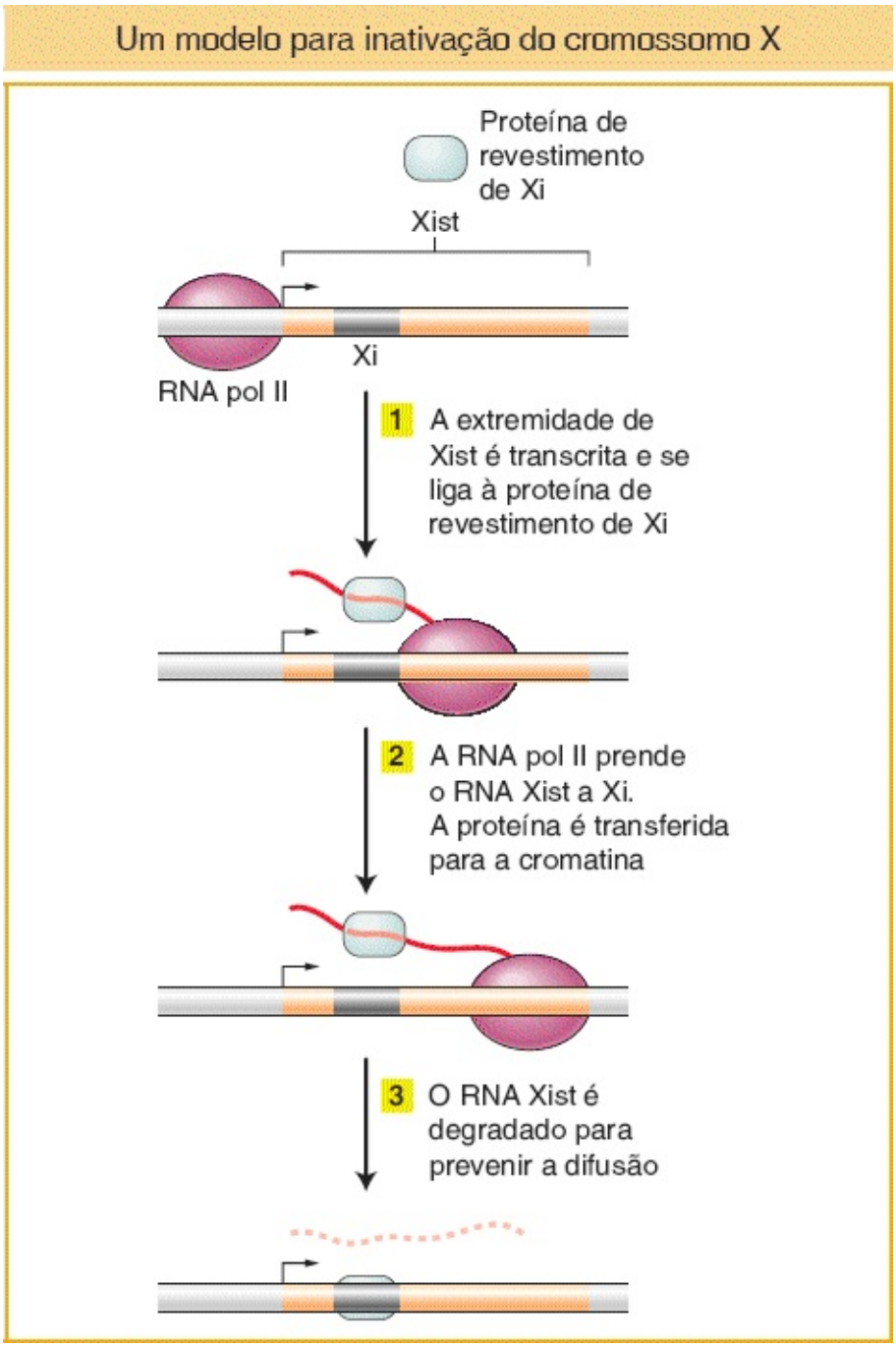


FIGURA 12.30 Um modelo que demonstra como o RNA Xist pode atuar em cis para ligar proteínas que inativam um cromossomo X por meio da formação de heterocromatina.

Dois tipos de pequenos RNA funcionais foram introduzidos no Capítulo 8: os siRNA e os miRNA. Nesta seção, exploramos como os miRNA auxiliam na regulação da expressão gênica eucariótica. A função dos siRNA será considerada

adicionalmente no Capítulo 15.

Lembre, do Capítulo 8, que os miRNA são sintetizados pela RNA pol II como RNA mais longos que são processados nos miRNA biologicamente ativos menores (aproximadamente 22 nt) (ver Figura 8.20). Os organismos contêm centenas de miRNA que regulam milhares de genes. Desses, aproximadamente 1/3 está organizado em grupos que são transcritos em um único transcrito, que posteriormente é processado para formar diversos miRNA. Contrariamente, aproximadamente 1/4 de todos os miRNA é processado a partir de transcritos derivados de íntrons recompostos. As etapas finais no processamento dos miRNA ocorrem no citoplasma.

No Capítulo 8, vimos como os miRNA de filamento único ativos se ligam ao complexo de silenciamento induzido por RNA (RISC) e hibridizam com mRNA complementares aos miRNA. Especificamente, a região de ligação do miRNA é composta pelos nucleotídios 2 a 8 do miRNA de aproximadamente 22 nt, denominada *região semente*. Os nucleotídios da região semente se ligam à 3′UTR de um mRNA que está sendo traduzido pelos ribossomos (Figura 12.31). Embora se saiba que o complexo miRNA-RISC inibe a tradução, o mecanismo preciso ainda está sob investigação. Modelos em relação a como a tradução pode ser reprimida incluem a interferência no início da tradução ou alongamento ou a remoção da cauda poli(A), que aceleraria a degradação do mRNA (Figura 12.31).

Embora os miRNA tenham sido descobertos há cerca de 20 anos, os cientistas estão apenas começando a decifrar a extensão e a complexidade do controle da expressão gênica eucariótica pelos miRNA. Considere que, em mamíferos, sequências complementares às regiões semente dos miRNA são encontradas na 3′UTR de diversas centenas de genes. Além disso, as 3′UTR de alguns genes contêm sequências complementares a diversos miRNA, enquanto muitos miRNA contêm sequências complementares às 3′UTR de diversos genes. Portanto, um gene possivelmente pode ser reprimido por diversos miRNA (seja individualmente ou em combinação) e um miRNA possivelmente pode reprimir a tradução de diversos genes. No Capítulo 11, você viu que a organização de genes bacterianos em óperons possibilitou a regulação coordenada de genes que contribuíram para um traço único, tal como a capacidade de utilizar o açúcar

lactose. Tendo em vista que a maior parte dos genes eucarióticos não está organizada em óperons, foi sugerido que a regulação pós-transcricional de diversos genes por um miRNA possibilita aos organismos superiores a capacidade de coordenar a expressão dos seus genes.

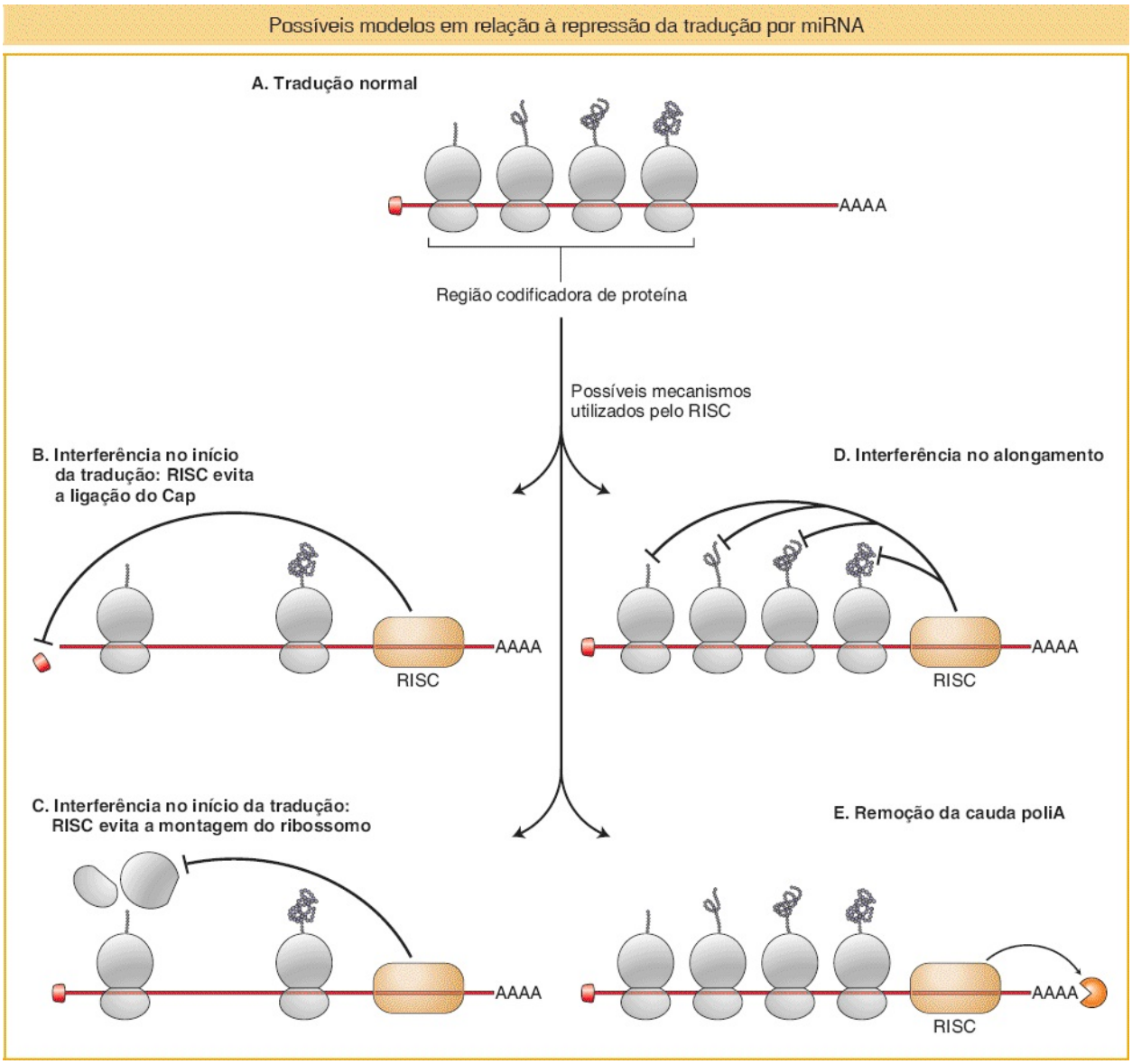


FIGURA 12.31 Ver o texto para detalhes.

RESUMO

Muitos aspectos da regulação gênica eucariótica se assemelham à regulação dos óperons bacterianos. Ambas operam amplamente no nível da transcrição e ambas dependem de proteínas transatuantes que se ligam às sequências-alvo reguladoras de ação cis na molécula de DNA. Essas proteínas reguladoras determinam o nível de transcrição de um gene ao controlar a ligação da RNA polimerase ao promotor gênico.

Existem três características de distinção importantes do controle da transcrição em eucariotos. Primeiramente, os genes eucarióticos apresentam acentuadores, que são elementos reguladores de ação cis localizados por vezes a grandes distâncias lineares do promotor. Muitos genes apresentam múltiplos acentuadores. Em segundo lugar, esses acentuadores com frequência são ligados por mais fatores de transcrição do que os óperons bacterianos. Eucariotos multicelulares devem gerar milhares de padrões de expressão gênica com um número limitado de proteínas reguladoras (fatores de transcrição). Eles realizam isso por meio de interações combinatórias de fatores de transcrição. Os acentuassomos são complexos de proteínas reguladoras que interagem de modo cooperativo e sinérgico para promover altos níveis de transcrição por meio do recrutamento da RNA polimerase II para o sítio de início da transcrição.

Em terceiro lugar, os genes eucarióticos estão compactados na cromatina. A ativação e a repressão dos genes necessitam de modificações específicas na cromatina. A vasta maioria dos milhares de genes em um genoma eucariótico típico está desligada em qualquer ocasião. Os genes são mantidos em um estado transcricionalmente inativo por meio da participação dos nucleossomos, que atuam para compactar a cromatina e prevenir a ligação da RNA polimerase II. A posição dos nucleossomos e a extensão da condensação da cromatina são determinadas pelo padrão de modificações pós-tradução das caudas de histonas. As modificações das histonas são marcas epigenéticas que, juntamente com a metilação das bases citosina, podem ser alteradas por fatores de transcrição. Esses fatores se ligam às regiões reguladoras e recrutam complexos proteicos que modificam enzimaticamente os nucleossomos adjacentes. Esses grandes complexos proteicos multissubunitários utilizam a energia da hidrólise de ATP para movimentar os nucleossomos e remodelar a cromatina.

A replicação do DNA copia fielmente ambas a sequência do DNA e a estrutura da cromatina da célula genitora para as células-filhas. Células recém-formadas herdam tanto a informação genética, inerente à sequência de nucleotídios do DNA, quanto a informação epigenética, que está no código de histonas e no padrão de metilação do DNA.

A existência de fenômenos epigenéticos, tais como o *imprinting* genético e a

inativação do cromossomo X, demonstra que a expressão gênica eucariótica pode ser silenciada sem a alteração da sequência de DNA do gene. Outro fenômeno epigenético, a variegação por efeito de posição, revelou a existência de domínios heterocromáticos repressivos que estão associados a nucleossomos altamente condensados e que contêm poucos genes. Isoladores de barreira mantêm a integridade do genoma ao evitar a conversão da eucromatina em heterocromatina.

Existe uma crescente apreciação em relação ao papel dos RNA funcionais, tais como os ncRNA e os miRNA, na regulação da expressão gênica eucariótica. Esses RNA direcionam complexos proteicos para o DNA complementar ou o RNA na célula. Em relação a alguns RNA (como o Xist), o ato de transcrição pode fixar o RNA a uma região cromossômica na qual proteínas se ligarão e alterarão a cromatina. Contrariamente, a tradução de centenas de mRNA é reprimida quando RISC ligado a miRNA complementares é direcionado às regiões 3′UTR dos mRNA.

TERMOS-CHAVE

acentuador
acentuassomo
cauda de histonas
coativador
código de histonas
compensação de dose
complexo mediador
corpúsculo de Barr
correpressor
domínio de ativação
efeito sinérgico
elemento proximal do promotor
eucromatina
gene repórter
hemimetilação

herança epigenética
heterocromatina
heterocromatina constitutiva
hiperacetilação
hipoacetilação
histona desacetilase (HDAC)
ilhota de CpG
imprinting genômico
imprinting materno
imprinting paterno
isolador de barreira
isolador de bloqueio de acentuador
marca epigenética
modificação das histonas
modificação pós-tradução (PTM)
remodelagem da cromatina
sequência de ativação *upstream* (UAS)
silenciamento gênico pós-transcricional
silenciamento gênico transcricional
variante de histonas
variegação por efeito de posição (PEV)

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 12.4, determinadas mutações diminuem a taxa de transcrição relativa do gene da β-globina. Onde essas mutações estão localizadas e como elas exercem seus efeitos sobre a transcrição?

2. Com base nas informações na Figura 12.6, como Gal4 regula quatro genes *GAL* diferentes ao mesmo tempo? Contraste esse mecanismo com o modo como o repressor Lac controla a expressão de três genes.

3. Em qualquer experimento, controles são essenciais para determinar o efeito específico da alteração de algum parâmetro. Na Figura 12.7, quais construtos são os "controles" que atuam para estabelecer o princípio de que os domínios de ativação são modulares e intercambiáveis?
4. Contraste o papel da proteína MCM1 nos diferentes tipos celulares de levedura demonstrados na Figura 12.10. Como os genes **a**-específicos são diferentemente controlados em tipos celulares diferentes?
5. Na Figura 12.11 B, em qual região cromossômica você provavelmente encontra a maior parte da proteína histona H1?
6. Qual é a conexão conceitual entre as Figuras 12.12 e 12.19?
7. Na Figura 12.19, onde o TATA boxe está localizado antes da formação do acentuassomo na parte superior da figura?
8. Digamos que você apresenta habilidades incríveis e consegue isolar as manchas brancas e vermelhas de tecido dos olhos de *Drosophila* demonstradas na Figura 12.24 com a finalidade de isolar o mRNA de cada preparação tecidual. Utilizando o seu conhecimento sobre as técnicas de DNA do Capítulo 10, desenhe um experimento que possibilitaria a você determinar se o RNA é transcrito a partir do gene branco no tecido vermelho ou no tecido branco ou ambos. Se for necessário, você tem acesso ao DNA radioativo do gene branco.
9. Na Figura 12.26, forneça um mecanismo bioquímico pelo qual a HP-1 consegue se ligar ao DNA apenas do lado esquerdo do isolador de barreira. De modo semelhante, por que a HMTase consegue se ligar apenas ao DNA à esquerda do isolador de barreira?
10. Em referência à Figura 12.28, desenhe o desfecho se houver uma mutação no gene B.

PROBLEMAS BÁSICOS

11. Quais analogias você consegue traçar entre os fatores de transcrição transatuantes que ativam a expressão gênica em eucariotos e os fatores

correspondentes em bactérias? Forneça um exemplo.

12. Contraste como os estados fundamentais dos genes no DNA de bactérias e eucariotos diferem em relação à ativação gênica (ver Figura 12.2).

13. Preveja e explique o efeito sobre a transcrição de *GAL1*, na presença somente de galactose, das mutações a seguir:

a. Deleção de um sítio de ligação de Gal4 no elemento UAS *GAL1*.

b. Deleção de todos os quatro sítios de ligação de Gal4 no elemento UAS *GAL1*.

c. Deleção do sítio de ligação Mig1 *upstream* de *GAL1*.

d. Deleção do domínio de ativação Gal4.

e. Deleção do gene *GAL80*.

f. Deleção do promotor *GAL1*.

g. Deleção do gene *GAL3*.

14. Como a ativação do gene *GAL1* é evitada na presença de galactose e glicose?

15. Quais são os papéis da desacetilação e da acetilação de histonas na regulação gênica, respectivamente?

16. Uma linhagem α de levedura que não consegue realizar a alteração do tipo reprodutivo é isolada. Quais mutações ela pode carregar que expliquem esse fenótipo?

17. Quais genes são regulados pelas proteínas α1 e α2 em uma célula α?

18. O que são as proteínas Sir? Como as mutações nos genes *SIR* afetam a expressão dos cassetes de tipo reprodutivo?

19. Qual é o significado do termo *herança epigenética*? Cite dois exemplos de tal herança?

20. O que é um acentuassomo? Por que uma mutação em qualquer uma das proteínas do acentuassomo poderia reduzir drasticamente a taxa de transcrição?

21. Normalmente, a deleção de um gene resulta em uma mutação recessiva. Explique como a deleção de um gene imprintado pode ser uma mutação

dominante.

22. O que é necessário ocorrer para que a expressão de dois genes diferentes em dois cromossomos diferentes seja regulada pelo mesmo miRNA?

23. Quais mecanismos se acredita serem responsáveis pela herança da informação epigenética?

24. Qual a diferença fundamental no modo como os genes bacterianos e eucarióticos são regulados?

25. Por que se diz que a regulação da transcrição em eucariotos é caracterizada por interações combinatórias?

26. Qual das declarações a seguir a respeito das histonas é verdadeira?

a. Elas são proteínas cuja sequência é altamente conservada em todos os eucariotos.

b. Elas são os blocos de construção dos nucleossomos.

c. a e **b** estão corretas.

d. Nenhuma das anteriores está correta.

27. Os nucleossomos são:

a. compostos por DNA e proteína.

b. necessários para condensar o DNA/os cromossomos.

c. essenciais para a correta regulação da expressão gênica eucariótica.

d. Todas as anteriores são verdadeiras.

28. As regiões cromossômicas que formam a heterocromatina:

a. contêm genes altamente expressos.

b. contêm poucos genes.

c. estão associadas à membrana nuclear.

d. são abundantes em procariotos.

29. A compensação de dose é necessária porque:

a. algumas regiões do genoma contêm mais genes do que outras.

b. os genes próximos da heterocromatina tendem a ser silenciados.

c. os acentuadores podem ativar a transcrição, estejam eles *upstream* ou *downstream* de um gene.

d. os genes no cromossomo X apresentam o dobro de cópias nas fêmeas do que nos machos.

30. Qual das seguintes é verdadeira a respeito da cromatina em procariotos?

a. Os cromossomos bacterianos não estão organizados em cromatina.

b. Está localizada no núcleo.

c. Está localizada no citoplasma.

d. É muito semelhante à cromatina em eucariotos.

31. Qual(is) das declarações a seguir é(são) verdadeira(s) sobre o fenótipo da cor de olhos variegada na Figura 12.23?

a. O gene branco está ativo nos setores brancos e inativo nos setores vermelhos.

b. O gene branco está inativo nos setores brancos e ativo nos setores vermelhos.

c. O gene branco está inativo nos setores brancos em virtude da difusão da heterocromatina.

d. b e **c** são, ambas, verdadeiras.

32. Quais dos seguintes são marcas epigenéticas?

a. Aminoácidos básicos metilados nas caudas de histonas.

b. Resíduos de citosina metilados no DNA.

c. Aminoácidos acetilados nas caudas de histonas.

d. Todas as anteriores são marcas epigenéticas.

PROBLEMAS DESAFIADORES

33. A transcrição de um gene denominado *YFG* (*your favorite gene*) é ativada quando três fatores de transcrição (TFA, TFB, TFC) interagem para recrutar o coativador CRX. TFA, TFB, TFC e CRX e seus respectivos sítios de ligação constituem um acentuassomo localizado a 10 kb do local de início da transcrição. Desenhe um diagrama demonstrando como você acredita que o acentuassomo atue para recrutar a RNA polimerase para o promotor de *YFG*.

34. Uma única mutação em um dos fatores de transcrição no Problema 33

resulta em uma redução drástica na transcrição de *YFG*. Diagrame a representação dessa interação mutante.

35. Diagrame o efeito de uma mutação no sítio de ligação para um dos fatores de transcrição no Problema 33.

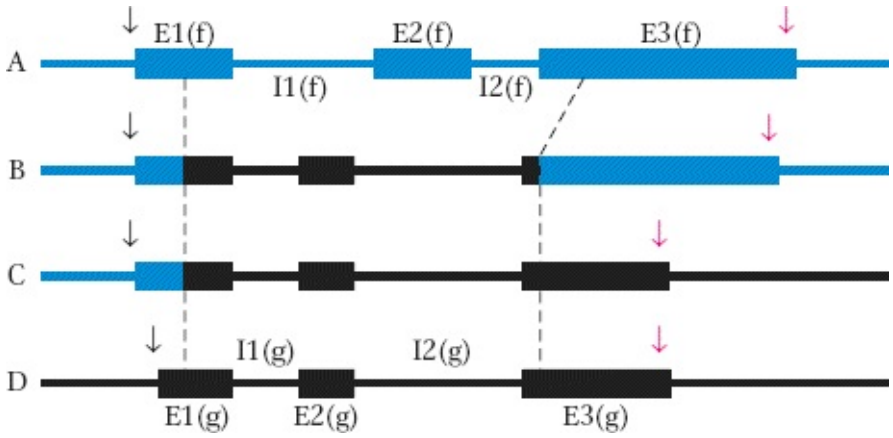
36. Alelos nulos (genes mutantes) não produzem produto proteico. Essa é uma alteração genética. Entretanto, genes epigeneticamente silenciados também não produzem produto proteico. Como se determina experimentalmente se um gene foi silenciado por mutação ou se foi silenciado epigeneticamente?

37. O que são marcas epigenéticas? Quais estão associadas à heterocromatina? Como se acredita que as marcas epigenéticas sejam responsáveis pela determinação da estrutura da cromatina?

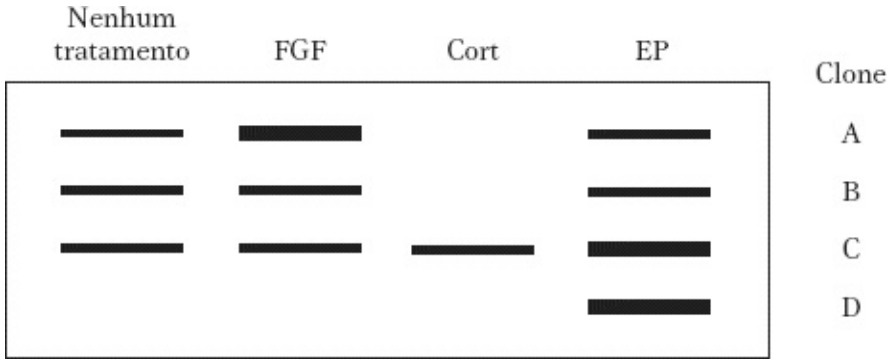
38. Você recebe quatro linhagens de leveduras pelo correio e as instruções de acompanhamento declaram que cada linhagem contém uma única cópia do transgene *A*. Você cultiva as quatro linhagens e determina que apenas três linhagens expressam o produto proteico do transgene *A*. Análise adicional revela que o transgene *A* está localizado em uma posição diferente no genoma de levedura em cada uma das quatro linhagens. Forneça uma hipótese para explicar esse resultado.

39. Você deseja encontrar os elementos reguladores de ação cis do DNA responsáveis pelas respostas de transcrição de dois genes, *c-fos* e *globina*. A transcrição do gene *c-fos* é ativada em resposta ao fator de crescimento de fibroblasto (FGF), mas é inibida pelo cortisol (Cort). Por outro lado, a transcrição do gene *globina* não é afetada pelo FGF ou pelo cortisol, mas é estimulada pelo hormônio eritropoetina (EP). Para encontrar os elementos reguladores do DNA de ação cis responsáveis por essas respostas transcricionais, você utiliza os clones a seguir dos genes *c-fos* e *globina*, bem como duas combinações "híbridas" (genes de fusão), conforme demonstrado no diagrama a seguir. A letra A representa o gene *c-fos* intacto, D representa o gene *globina* intacto e B e C representam fusões dos genes *c-fos* e *globina*. Os éxons (E) e os íntrons (I) de *c-fos* e *globina* estão numerados. Por exemplo, E3(f) é o terceiro éxon do gene *c-fos* e I2(g) é o

segundo íntron do gene *globina*. (Esses rótulos são fornecidos para ajudar a tornar a sua resposta clara.) Os sítios de início da transcrição (*setas pretas*) e os sítios de poliadenilação (*setas vermelhas*) estão indicados.



Você introduz todos os quatro clones simultaneamente em células de cultura de tecido e em seguida estimula alíquotas individuais dessas células com um dos três fatores. A análise em gel do RNA isolado a partir das células fornece os resultados a seguir:



Os níveis de transcritos produzidos pelos genes introduzidos em resposta a diversos tratamentos estão demonstrados; a intensidade dessas bandas é proporcional à quantidade de transcritos produzidos a partir de um clone em particular. (A falha na visualização de uma banda indica que o nível de transcrito é indetectável.)

a. Onde está o elemento de DNA que possibilita a ativação por FGF?

b. Onde está o elemento de DNA que possibilita a repressão por Cort?

c. Onde está o elemento de DNA que possibilita a indução por EP? Explique a sua resposta.

40. Com a utilização do sistema experimental demonstrado na Figura 12.26, um geneticista é capaz de inserir um isolador de barreira entre o gene *white* e a heterocromatina. O fenótipo do olho de *Drosophila* transgênica mais provavelmente seria:

a. todo branco, tendo em vista que o elemento de barreira evitará a difusão da heterocromatina.

b. todo vermelho, tendo em vista que o elemento de barreira evitará a difusão da heterocromatina.

c. ainda variegado, tendo em vista que os elementos de barreira não interrompem a difusão da heterocromatina.

d. ainda variegado, tendo em vista que os elementos de barreira não atuam em *Drosophila*.

41. Qual dos seguintes é um exemplo de repressão gênica pós-transcricional?

a. A quantidade de RNA transcrito a partir de um gene é reduzida em virtude da metilação do DNA.

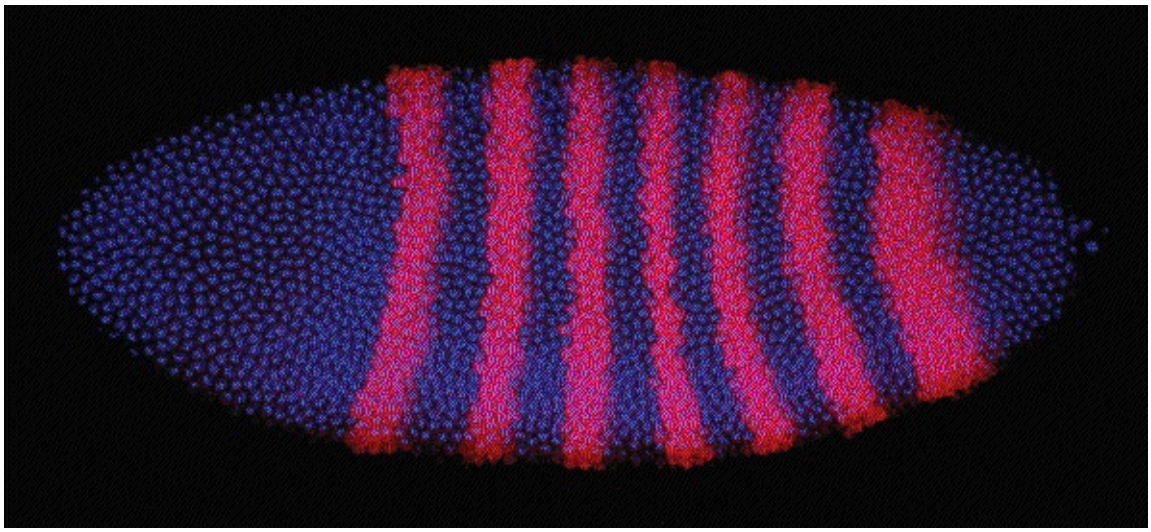
b. A quantidade de RNA é reduzida em virtude de sua rápida degradação.

c. A quantidade de proteína produzida é reduzida por meio da ação de um miRNA.

d. As alternativas b e c estão ambas corretas.

CAPÍTULO 13

Controle Genético do Desenvolvimento



Expressão gênica em um embrião de mosca-das-frutas em desenvolvimento. As sete listras *magenta* marcam as células que expressam o mRNA de um gene que codifica uma proteína reguladora que controla o número de segmentos no embrião de *Drosophila*. A regulação espacial da expressão gênica é central para o controle do desenvolvimento animal. (Dave Kosman, Ethan Bier e Bill McGinnis.)

TÓPICOS

13.1 Abordagem genética do desenvolvimento
13.2 Toolkit genética para o desenvolvimento de Drosophila
13.3 Definição da toolkit inteira
13.4 Regulação espacial da expressão gênica no desenvolvimento
13.5 Regulação pós-transcricional da expressão gênica no desenvolvimento
13.6 De moscas a dedos, penas e placas do assoalho | Os muitos papéis dos genes toolkit
13.7 Desenvolvimento e doença

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Distinguir os membros da *toolkit* genética para o desenvolvimento de outros genes e explicar como eles são identificados
- Correlacionar onde e quando os genes reguladores de padrões são expressos durante o desenvolvimento com os fenótipos que resultam de mutações neles
- Explicar exemplos de como padrões de expressão gênica espacialmente restritos são gerados durante o desenvolvimento
- Relacionar as funções bioquímicas das proteínas toolkit aos seus efeitos sobre o desenvolvimento dos corpos ou das partes corporais
- Identificar componentes homólogos da *toolkit* genética em diferentes filos de animais.

De todos os fenômenos na biologia, poucos inspiram mais admiração do que a formação de um animal complexo a partir de uma célula única, o zigoto. Nessa transformação espetacular, forças invisíveis organizam a massa de células em divisão em uma forma com uma cabeça e uma cauda distintas, vários apêndices e muitos órgãos. O grande geneticista Thomas Hunt Morgan não ficou imune a esse apelo estético:

> Um zigoto transparente, na medida em que se desenvolve, é um dos objetos mais fascinantes no mundo dos seres vivos. A contínua alteração na forma que ocorre hora a hora nos intriga em virtude de sua grande simplicidade. Os padrões geométricos que se apresentam a cada mudança convidam à análise matemática [...] Esse espetáculo faz um apelo irresistível aos lados emocional e artístico de nossa natureza.[1]

Entretanto, por toda a sua beleza e fascinação, biólogos foram confundidos durante décadas a respeito de como a forma biológica é gerada durante o desenvolvimento. Morgan também disse que “se o mistério que circunda a embriologia sequer chegar a ser da nossa compreensão, devemos [...] recorrer a outros meios além da descrição do espetáculo que se apresenta”.

A longa estagnação na embriologia durou muito mais do que a época de Morgan nas décadas de 1910 e 1920, mas finalmente foi rompida por geneticistas

que trabalhavam muito na tradição da genética ao estilo de Morgan e com seu modelo genético favorito e mais produtivo, a mosca-das-frutas *Drosophila melanogaster*.

Os catalisadores-chave para a compreensão da formação dos animais foram as descobertas de monstros genéticos — moscas-das-frutas mutantes com alterações dramáticas das estruturas corporais (Figura 13.1). Nos primórdios da genética de *Drosophila*, mutantes raros com transformações espetaculares de partes do corpo surgiram espontaneamente ou como subprodutos de outros experimentos. Em 1915, Calvin Bridges, então aluno de Morgan, isolou uma mosca com uma mutação que fazia com que as pequenas asas traseiras (halteres) da mosca-das-frutas se assemelhassem às grandes asas dianteiras. Ele apelidou o mutante de *bithorax*. A transformação em mutantes *bithorax* é denominada *homeótica* (do grego *homeos*, que significa o mesmo ou semelhante), tendo em vista que uma parte do corpo (a asa traseira) é transformada e se assemelha a outra (a asa dianteira), conforme demonstrado na Figura 13.1 B. Subsequentemente, diversos outros mutantes homeóticos foram identificados em *Drosophila*, tais como o dramático mutante *Antennapedia*, no qual pernas se desenvolvem no lugar das antenas (Figura 13.1 C).

Os efeitos espetaculares dos mutantes homeóticos inspiraram o que se tornaria uma revolução na embriologia, após terem sido disponibilizados recursos de biologia molecular para compreender o que os genes homeóticos codificavam e como exerciam tamanha influência no desenvolvimento de partes corporais inteiras. Surpreendentemente, esses estranhos genes da mosca-das-frutas se tornaram um passaporte para o estudo de todo o reino animal, na medida em que foram descobertos correspondentes desses genes que desempenhavam papéis semelhantes em quase todos os animais.

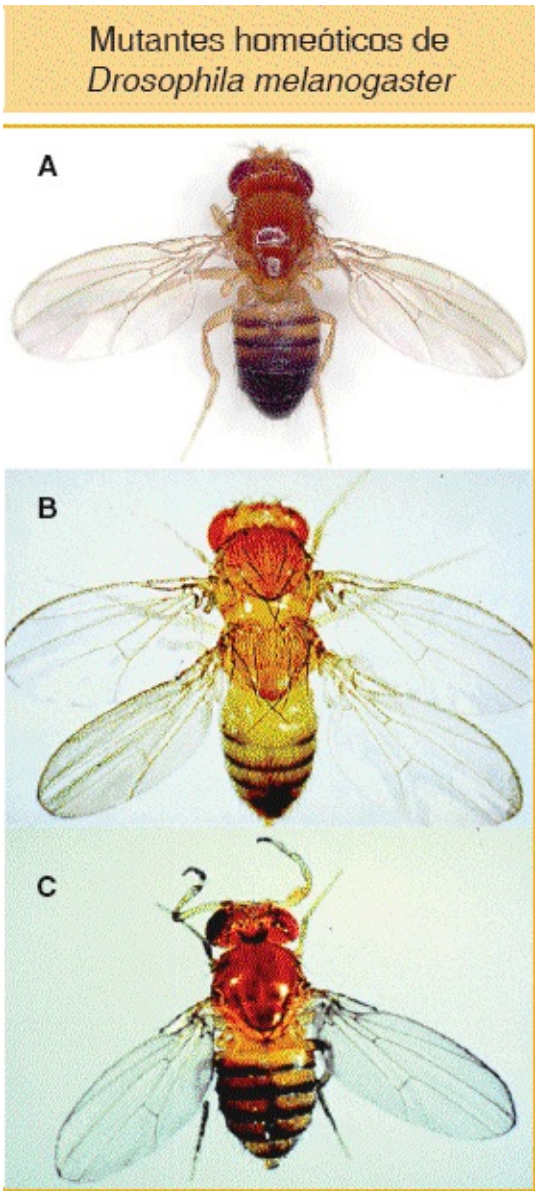


FIGURA 13.1 Em mutantes homeóticos, a identidade de uma estrutura corporal foi alterada para outra. **A.** Mosca normal com um par de asas dianteiras no segundo segmento torácico e um par de pequenas asas traseiras no terceiro segmento torácico. **B.** Mutante triplo para três mutações no gene *Ultrabithorax*. A função de *Ubx* é perdida na parte posterior do tórax, o que causa o desenvolvimento de asas dianteiras no lugar das asas traseiras. **C.** Mutante *Antennapedia*, no qual as antenas são transformadas em pernas. (Sean Carroll.)

O estudo do desenvolvimento de animais e plantas é uma disciplina muito ampla e ainda em crescimento. Como tal, não tentamos uma visão geral abrangente da embriologia. Em vez disso, neste capítulo, enfocaremos alguns poucos conceitos gerais que ilustram a lógica do controle genético do

desenvolvimento de animais. Exploraremos como a informação para a construção de estruturas complexas é codificada no genoma. Contrariamente ao controle da regulação gênica em células únicas bacterianas ou eucarióticas, o controle genético da formação e da padronização corporal é fundamentalmente uma questão de regulação gênica no *espaço* tridimensional e no *tempo*. Veremos, ainda, que os princípios que regem o controle genético do desenvolvimento estão conectados àqueles já apresentados nos Capítulos 11 e 12, determinantes do controle fisiológico da expressão gênica em bactérias e eucariotos unicelulares.

13.1 Abordagem genética do desenvolvimento

Durante muitas décadas, o estudo do desenvolvimento embrionário envolveu sobretudo a manipulação física de embriões, células e tecidos. Foram estabelecidos diversos conceitos-chave a respeito das propriedades de embriões em desenvolvimento por meio de experimentos nos quais parte de um embrião foi transplantada em outra. Por exemplo, o transplante de uma parte de um embrião de anfíbio em desenvolvimento para outro local em um embrião receptor induz o tecido adjacente a formar um segundo eixo corporal completo (Figura 13.2 A). De modo semelhante, o transplante da parte posterior de um membro de pintinho em desenvolvimento para a anterior pôde induzir dedos extras, mas com polaridade reversa em relação aos dedos normais (Figura 13.2 B). Essas regiões transplantadas do embrião de anfíbio e do membro de galinha em desenvolvimento foram denominadas *organizadores*, em virtude de sua extraordinária capacidade de organizar o desenvolvimento de tecidos adjacentes. Foi postulado que as células nos organizadores produzem *morfógenos*, moléculas que induziam diversas respostas no tecido adjacente de modo dependente da concentração.

Embora esses resultados experimentais fossem espetaculares e fascinantes, não houve progresso adicional na compreensão da natureza dos organizadores e dos morfógenos após a sua descoberta na primeira metade da década de 1900. Era praticamente impossível isolar as moléculas responsáveis por essas atividades

com a utilização de técnicas de separação bioquímica. As células embrionárias produzem milhares de substâncias — proteínas, glicolipídios, hormônios e assim por diante. Um morfógeno pode ser qualquer uma dessas moléculas, mas estaria presente em quantidades minúsculas — uma agulha no palheiro dos produtos celulares.

O longo impasse na definição da embriologia em termos moleculares foi rompido por abordagens genéticas — principalmente o isolamento sistemático de mutantes com defeitos discretos no desenvolvimento e a subsequente caracterização e estudo dos produtos gênicos que eles codificavam. A abordagem genética para o estudo do desenvolvimento apresentou muitas vantagens sobre as estratégias bioquímicas alternativas. Primeiramente, o geneticista não precisa realizar quaisquer presunções a respeito do número ou da natureza das moléculas necessárias para um processo. Em segundo lugar, a quantidade (limitada) de um produto gênico não é um impedimento: todos os genes podem sofrer mutação, independentemente da quantidade de produto produzido por um gene. Em terceiro lugar, a abordagem genética pode revelar fenômenos para os quais não existe ensaio bioquímico ou outro bioensaio.

Do ponto de vista genético, existem quatro questões-chave a respeito do número, da identidade e da função dos genes que participam no desenvolvimento:

1. Quais genes são importantes no desenvolvimento?
2. Onde no animal em desenvolvimento e em quais ocasiões esses genes estão ativos?
3. Como a expressão dos genes do desenvolvimento é regulada?
4. Por meio de quais mecanismos moleculares os produtos gênicos afetam o desenvolvimento?

Para abordar essas questões, precisaram ser planejadas estratégias para identificar, catalogar e analisar os genes que controlam o desenvolvimento. Uma das primeiras considerações na análise genética do desenvolvimento animal foi qual animal deveria ser estudado. Dentre os milhões de espécies vivas, quais eram as mais promissoras? A mosca-das-frutas *Drosophila melanogaster* surgiu

como o modelo genético líder do desenvolvimento animal, em virtude de sua facilidade de criação, do rápido ciclo de vida, da citogenética e das décadas de análise genética clássica (incluindo o isolamento de muitos mutantes notáveis), que proporcionavam vantagens experimentais importantes (ver Organismo-modelo, *Drosophila*). O nematódeo *Caenorhabditis elegans* também apresentava diversas características atraentes, mais particularmente sua estrutura simples e linhagens celulares bem-estudadas. Entre os vertebrados, o desenvolvimento das técnicas de alteração de genes-alvo disponibilizou o camundongo de laboratório para estudos genéticos mais sistemáticos e o peixe-zebra *Danio rerio* recentemente se tornou um modelo favorito em virtude da transparência do embrião e dos avanços no seu estudo genético. Entre as plantas, a *Arabidopsis thaliana* tem desempenhado um papel semelhante ao da *Drosophila* no esclarecimento dos mecanismos fundamentais no desenvolvimento das plantas.

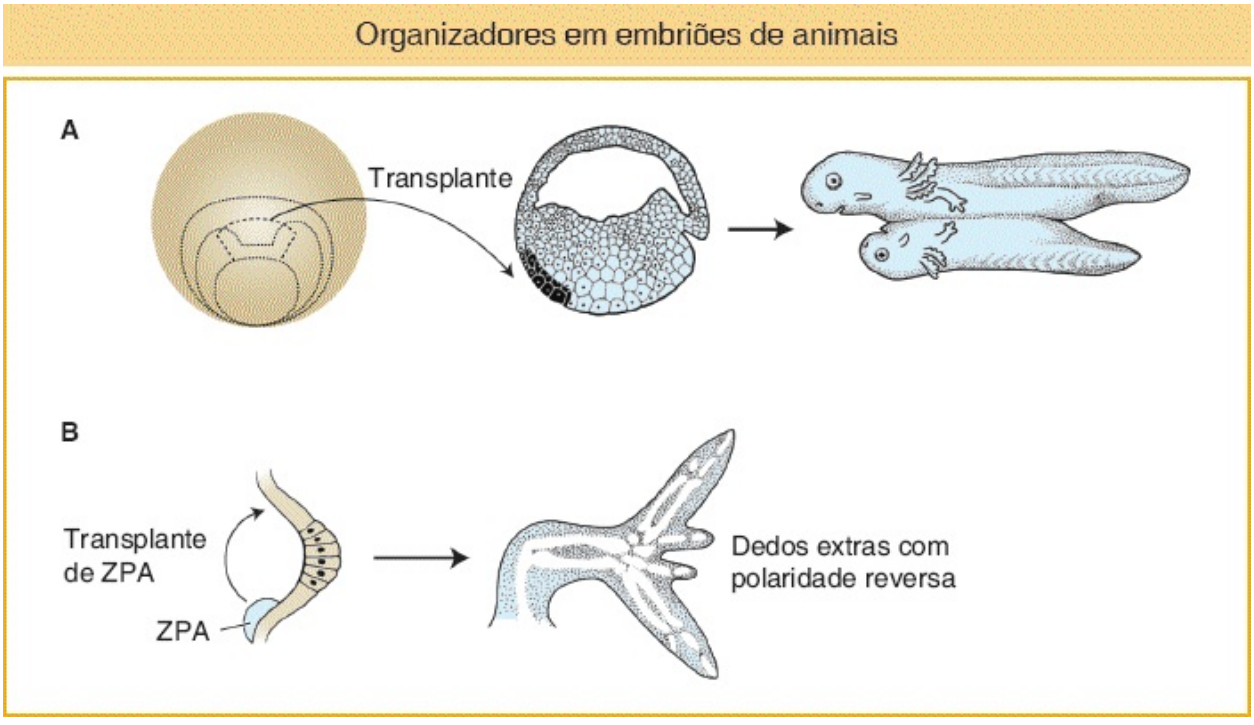


FIGURA 13.2 Experimentos com transplantes desempenharam um papel central no início da embriologia e demonstraram a atividade de organização de longo alcance dos tecidos embrionários. **A.** O organizador Spemann. O "lábio" do blastóporo dorsal de um embrião de anfíbio inicial pode induzir um segundo eixo embrionário e um novo embrião quando transplantado para a região ventral de um embrião receptor. **B.** No broto do membro vertebral de galinha em desenvolvimento, a zona de atividade polarizada (ZPA) organiza o

padrão ao longo do eixo anteroposterior. O transplante de ZPA de uma posição posterior para uma anterior induz dedos extras com polaridade reversa.

ORGANISMO-MODELO *Drosophila*

Análise mutacional do desenvolvimento inicial de Drosophila

As percepções iniciais sobre o controle genético da formação de padrões tiveram origem a partir de estudos da mosca-das-frutas *Drosophila melanogaster*. O desenvolvimento da *Drosophila* comprovou ser uma mina de ouro para os pesquisadores, tendo em vista que os problemas do desenvolvimento podem ser abordados por meio da utilização de técnicas genéticas e moleculares simultaneamente.

O embrião de *Drosophila* tem sido especialmente importante na compreensão da formação do plano corporal animal básico. Um motivo para isso é que uma anormalidade no plano corporal de um mutante é facilmente identificada no exoesqueleto larval do embrião da *Drosophila*. O exoesqueleto larval é uma estrutura não celular, composta por um polímero de polissacarídio denominado quitina, que é produzido como uma secreção das células epidérmicas do embrião. Cada estrutura do exoesqueleto é formada a partir de células epidérmicas ou imediatamente adjacentes àquela estrutura. Com seu padrão intricado de cerdas, endentações e outras estruturas, o exoesqueleto proporciona diversos pontos de referência que atuam como indicadores dos destinos atribuídos às muitas células epidérmicas. Em particular, existem muitas estruturas anatômicas distintas ao longo dos eixos anteroposterior (A-P) e dorsoventral (D-V) (ver adiante). Além disso, tendo em vista que todos os nutrientes necessários para o desenvolvimento do estágio larval estão pré-acondicionados no zigoto, embriões mutantes nos quais os destinos das células A-P ou D-V são drasticamente alterados podem, não obstante, desenvolver-se até o final da embriogênese e produzir uma larva mutante em aproximadamente 1 dia (ver ilustração). O exoesqueleto de uma referida larva mutante espelha os destinos mutantes atribuídos aos subconjuntos das células epidérmicas e, assim, pode identificar genes valiosos

para uma análise detalhada.

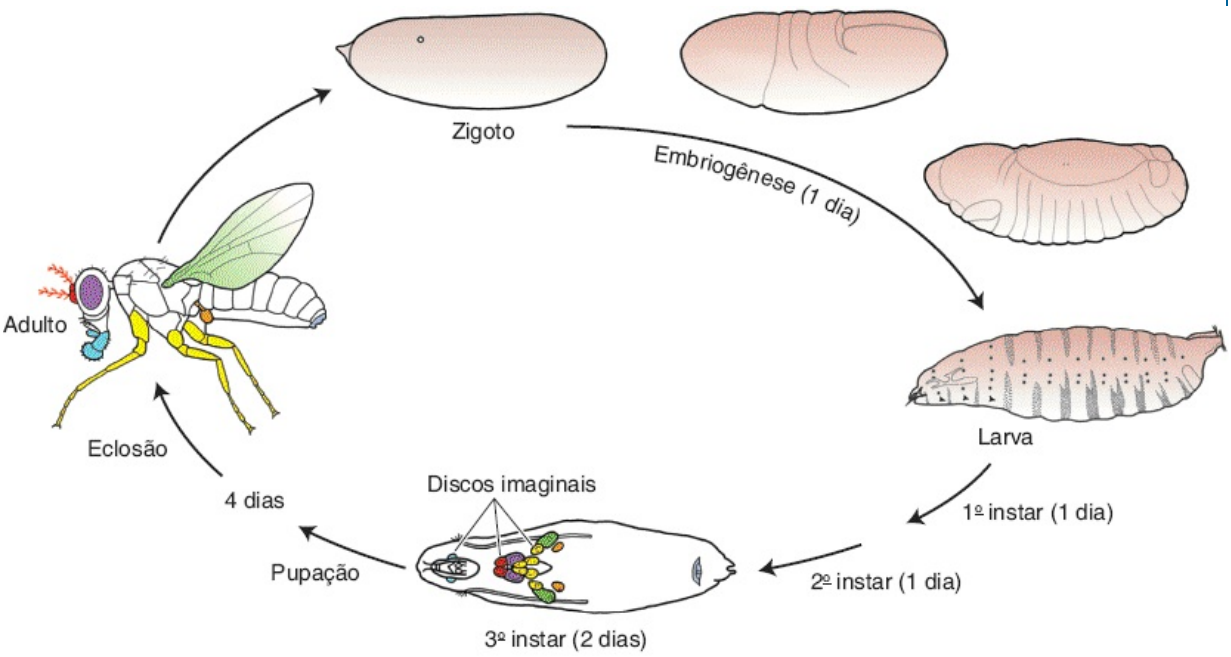
O desenvolvimento do padrão corporal adulto da *Drosophila* demora pouco mais de 1 semana (ver ilustrações). Pequenas populações de células reservadas durante a embriogênese proliferam durante três estágios larvais (instars) e se diferenciam no estágio de pupa em estruturas adultas. Essas células reservadas incluem os *discos imaginais*, regiões com formato de disco que dão origem a apêndices e tecidos específicos em cada segmento, como pernas, asas, olhos e discos de antenas. Os discos imaginais são de fácil remoção para a análise da expressão gênica (ver Figura 13.7).

Os genes que contribuem para o plano corporal de *Drosophila* podem ser clonados e caracterizados no nível molecular com facilidade. A análise dos genes clonados com frequência fornece informações valiosas sobre a função do produto proteico — normalmente pela identificação de correlatos próximos na sequência de aminoácidos do polipeptídio codificado por meio de comparações com todas as sequências proteicas armazenadas em bases de dados públicas. Além disso, pode-se investigar os padrões espacial e temporal da expressão (1) de um mRNA, por meio da utilização de sequências de DNA unifilamentar marcadas histoquimicamente complementares ao mRNA para a realização da hibridização *in situ* do RNA ou (2) de uma proteína, por meio da utilização de anticorpos marcados histoquimicamente que se ligam especificamente à proteína.

Utilização do conhecimento de um organismo-modelo para acelerar a descoberta de genes de desenvolvimento em outros organismos

Com a descoberta de que existem diversos genes *homeoboxe* no genoma de *Drosophila*, as semelhanças entre as sequências de DNA desses genes podem ser exploradas em caças ao tesouro em relação a outros membros da família do gene homeótico. Essas caçadas dependem da complementaridade dos pares de bases do DNA. Para essa finalidade, foram realizadas hibridizações de DNA sob *condições de estringência moderada*, nas quais pode haver algum

pareamento incorreto de bases entre os filamentos hibridizados sem prejudicar o pareamento adequado por ligações de hidrogênio dos pares de bases próximos. Algumas dessas caças ao tesouro foram realizadas no próprio genoma de *Drosophila*, na busca por mais membros da família. Outras procuraram genes *homeoboxe* em outros animais, por meio de *zoo blots* (*Southern blots* de DNA digerido por enzima de restrição de diferentes animais), por meio da utilização do DNA *homeoboxe* radioativo de *Drosophila* como sonda. Essa abordagem levou à descoberta de sequências *homeoboxe* homólogas em muitos animais diferentes, incluindo seres humanos e camundongos. (De fato, essa é uma abordagem muito poderosa para "pescar" correlatos de quase todos os genes em seu organismo favorito.) Agora os genes homólogos são tipicamente identificados por meio de pesquisas computadorizadas de sequências genômicas (Capítulo 14).



Visão geral do desenvolvimento de *Drosophila*. A larva se forma em 1 dia e em seguida é submetida a diversos estágios de crescimento, durante os quais os discos imaginais e outros precursores de estruturas adultas proliferam. Estas estruturas se diferenciam durante a pupação e a mosca adulta eclode (eclosão) e inicia o ciclo novamente

Por meio de análise genética sistemática e direcionada, bem como de estudos genômicos comparativos, foi definida uma grande parte da **caixa de ferramentas genéticas** (*toolkit*) para o desenvolvimento dos corpos, das partes corporais e dos tipos celulares de diferentes espécies animais. Primeiramente, enfocaremos a *toolkit* genética de *Drosophila melanogaster*, tendo em vista que a sua identificação foi uma fonte de percepções importantes sobre o controle genético do desenvolvimento; a sua descoberta catalisou a identificação da *toolkit* genética de outros animais, incluindo seres humanos.

13.2 Toolkit genética para o desenvolvimento de Drosophila

Os genomas animais tipicamente contêm aproximadamente 13.000 a 22.000 genes. Muitos desses genes codificam proteínas que atuam em processos essenciais em todas as células do corpo (p. ex., no metabolismo celular ou na biossíntese de macromoléculas). Os referidos genes com frequência são denominados **genes de manutenção**. Outros genes codificam proteínas que realizam as tarefas especializadas de diversos sistemas orgânicos, tecidos e células do corpo, tais como as proteínas globinas no transporte de oxigênio ou proteínas de anticorpos que medeiam a imunidade. Aqui, estamos interessados em um conjunto diferente de genes, aqueles que dizem respeito à formação de órgãos e tecidos e à especificação dos tipos celulares — a *toolkit* genética para o desenvolvimento que determina o plano corporal geral e o número, a identidade e o padrão das partes corporais.

Os genes *toolkit* da mosca-das-frutas em geral foram identificados por meio das monstruosidades ou catástrofes que surgem quando eles estão mutados. Mutações nos genes *toolkit* de duas fontes produziram a maior parte do nosso conhecimento. A primeira fonte é composta por mutações espontâneas que surgem em populações de laboratório. A segunda, por mutações induzidas aleatoriamente

por meio do tratamento com mutágenos (tais como substâncias químicas ou radiação) que aumentam muito a frequência de genes danificados em todo o genoma. Os elegantes refinamentos da última abordagem tornaram possíveis pesquisas sistemáticas por mutantes que identificaram muitos membros da *toolkit* genética da mosca. Os membros dessa *toolkit* constituem apenas uma pequena fração, talvez diversas centenas de genes, dos aproximadamente 14.000 genes no genoma da mosca.

> CONCEITO-CHAVE A *toolkit* genética para o desenvolvimento animal é composta por uma pequena fração de todos os genes. Apenas um pequeno subconjunto de todo o complemento de genes no genoma afeta o desenvolvimento de modos bem-definidos.

Classificação dos genes pela função no desenvolvimento

Uma das primeiras tarefas após a execução de uma triagem genética em relação a mutações é classificar aquelas de interesse. Muitas mutações são letais quando hemizigotas ou homozigotas, tendo em vista que as células não conseguem sobreviver sem os produtos afetados por essas mutações. As mutações mais interessantes são aquelas que causam algum defeito bem-definido sobre o padrão corporal embrionário ou adulto ou ambos. Comprovou-se útil agrupar os genes afetados por mutações em diversas categorias com base na natureza de seus fenótipos mutantes. Muitos genes *toolkit* podem ser classificados de acordo com a sua função no controle da identidade das partes corporais (p. ex., de diferentes segmentos ou apêndices), na formação de partes corporais (p. ex., de órgãos e anexos), no número de partes corporais, na formação de tipos celulares e na organização dos eixos corporais primários (os eixos anteroposterior, ou A-P, e dorsoventral, ou D-V).

Iniciaremos o nosso inventário da *toolkit* de *Drosophila* por meio do exame dos genes que controlam a identidade dos segmentos e dos apêndices. Fazemos isso para fins históricos e conceituais. Os genes que controlam a identidade dos

segmentos e dos apêndices estão entre os primeiros genes *toolkit* identificados. As descobertas subsequentes a respeito da sua natureza foram fontes de percepções profundas não apenas sobre como os seus produtos atuam, mas também sobre o conteúdo e o funcionamento da *toolkit* da maior parte dos animais. Além disso, seus espetaculares fenótipos mutantes indicam que eles estão entre os genes de ação mais global que afetam a forma animal. O aprendizado a respeito desses genes deve estimular os nossos apetites para aprender mais a respeito de toda a *toolkit* que controla o desenvolvimento da forma animal.

Genes homeóticos e identidade segmentar

Entre as anormalidades mais fascinantes a serem descritas em animais estão aquelas nas quais uma parte corporal normal é substituída por outra. As referidas transformações homeóticas foram observadas em muitas espécies na natureza, incluindo dípteros, nos quais uma perna forma-se no lugar de uma antena, e rãs, nas quais uma vértebra torácica forma-se no lugar de uma vértebra cervical (Figura 13.3). Embora comumente apenas um membro de um par de estruturas bilaterais esteja alterado em muitas variantes de ocorrência natural, ambos os membros de um par de estruturas bilaterais estão alterados em mutantes homeóticos das moscas-das-frutas. No caso anterior, a alteração não é hereditária, mas mutantes homeóticos reproduzem-se de geração para geração.

Em uma transformação homeótica, uma parte do corpo é substituída por outra

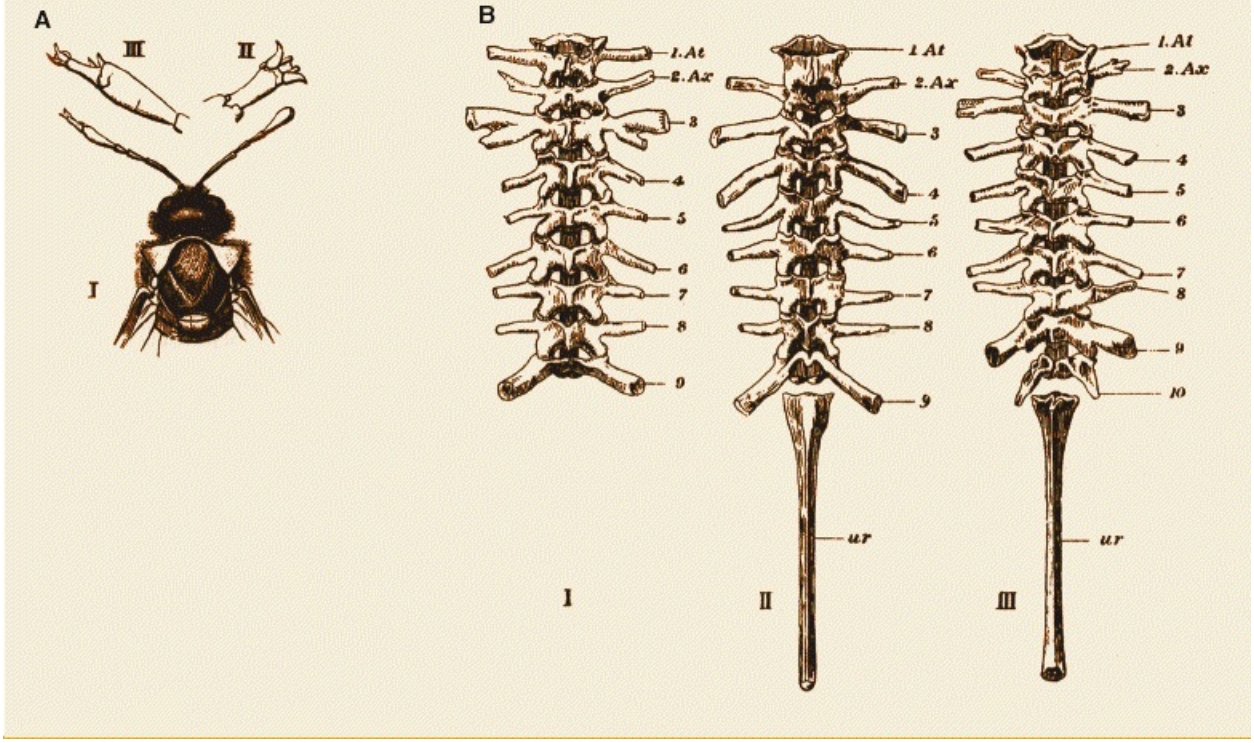


FIGURA 13.3 Um desenho do fim do século 19 de um dos primeiros estudos de transformações homeóticas na natureza. **A.** Homeose em um díptero, com a antena esquerda transformada em uma perna. **B.** Homeose em uma rã. A amostra *intermediária* é normal. A amostra *à esquerda* apresenta estruturas extras que crescem para fora da parte superior da coluna vertebral. A amostra *à direita* apresenta um par de vértebras extra. (De W. Bateson, *Material for the Study of Variation*, Macmillan, 1894.)

A fascinação científica pelos mutantes homeóticos tem origem em três propriedades. Primeiramente, é impressionante que uma única mutação gênica possa alterar uma via do desenvolvimento tão dramaticamente. Em segundo lugar, é surpreendente que a estrutura formada no mutante seja um similar bem-desenvolvido de outra parte corporal. Por fim, é importante observar que mutações homeóticas transformam a identidade de **estruturas serialmente reiteradas**. Os corpos de insetos e de muitos animais são compostos por partes repetidas de estruturas semelhantes, como blocos estruturais, arranjados em uma série. As asas dianteiras e traseiras, os segmentos e as antenas, as pernas e as partes bucais dos insetos são conjuntos de partes corporais serialmente

reiteradas. As mutações homeóticas transformam identidades dentro desses grupos.

Uma mutação pode causar uma perda de função do gene homeótico onde o gene normalmente atua ou pode causar um ganho de função homeótica onde o gene homeótico normalmente não atua. Por exemplo, o gene *Ultrabithorax* (*Ubx*) atua no desenvolvimento da asa traseira, promovendo o desenvolvimento da asa traseira e reprimindo o da asa dianteira. Mutações de perda de função em *Ubx* transformam a asa traseira em dianteira. Mutações de ganho de função dominantes em *Ubx* transformam a asa dianteira em traseira. De modo semelhante, as transformações de antena em perna dos mutantes *Antennapedia* (*Antp*) são causadas pelo ganho dominante de função de *Antp* na antena. Além dessas transformações na identidade dos apêndices, as mutações homeóticas podem transformar a identidade de segmentos, fazendo com que um segmento corporal do adulto ou da larva se assemelhe a outro.

Embora os genes homeóticos tenham sido identificados pela primeira vez por meio de mutações espontâneas que afetam moscas adultas, eles são necessários durante a maior parte do desenvolvimento de uma mosca. Pesquisas sistemáticas em relação a genes homeóticos levaram à identificação de oito *loci*, atualmente denominados **genes *Hox***, que afetam a identidade dos segmentos e de seus apêndices correlatos em *Drosophila*. Em geral, a perda completa de qualquer função dos genes *Hox* é letal no início do desenvolvimento. As mutações dominantes que transformam adultos são viáveis nos heterozigotos, tendo em vista que o alelo do tipo selvagem proporciona a função gênica normal no animal em desenvolvimento.

Organização e expressão dos genes Hox

Uma característica muito intrigante dos genes *Hox* é que eles estão agrupados em dois **complexos gênicos** que estão localizados no terceiro cromossomo de *Drosophila*. O complexo *Bithorax* contém três genes *Hox* e o complexo *Antennapedia*, cinco. Além disso, a ordem dos genes nos complexos e no cromossomo corresponde à ordem das regiões corporais, da cabeça à cauda, que são influenciadas por cada gene *Hox*.

A relação entre a estrutura dos complexos de genes *Hox* e os fenótipos dos mutantes de genes *Hox* foi esclarecida pela caracterização molecular dos genes. A clonagem molecular das sequências que compreendem cada *locus Hox* forneceu os meios para analisar onde no animal em desenvolvimento cada gene é expresso. Esses aspectos espaciais da expressão e da regulação gênica são cruciais para compreender a lógica do controle genético do desenvolvimento. Em relação aos genes *Hox* e a outros genes *toolkit*, o desenvolvimento de tecnologias que tornaram possível a visualização da expressão gênica e proteica foi crucial para a compreensão da relação entre a organização gênica, a função gênica e os fenótipos mutantes.

Duas tecnologias principais para a visualização da expressão gênica em embriões ou outros tecidos são (1) a expressão de transcritos de RNA visualizada por meio de hibridização *in situ* e (2) a expressão de proteínas visualizadas por meio de métodos imunológicos. Cada tecnologia depende do isolamento de clones de cDNA que representam o transcrito de mRNA maduro e a proteína (Figura 13.4).

No embrião em desenvolvimento, os genes *Hox* são expressos em domínios espacialmente restritos, por vezes sobrepostos, dentro do embrião (Figura 13.5). Os genes também são expressos nos tecidos da larva e da pupa que darão origem às partes corporais do adulto.

Os padrões de expressão dos genes *Hox* (e outros genes *toolkit*) em geral estão correlacionados com as regiões do animal afetadas pelas mutações gênicas. Por exemplo, o sombreamento azul-escuro na Figura 13.5 indica onde o gene *Ubx* é expresso. Esse gene *Hox* é expresso no segmento torácico posterior e na maior parte dos segmentos abdominais do embrião. O desenvolvimento desses segmentos está alterado em mutantes *Ubx*. O gene *Ubx* também é expresso nas asas traseiras em desenvolvimento, mas não nas asas dianteiras em desenvolvimento (Figura 13.6), conforme seria esperado sabendo-se que o *Ubx* promove o desenvolvimento das asas traseiras e reprime o desenvolvimento das asas dianteiras nesse apêndice.

> CONCEITO-CHAVE A expressão espacial dos genes *toolkit* normalmente está correlacionada de modo próximo com as regiões do animal afetadas por mutações gênicas.

É crucial distinguir o papel dos genes *Hox* na determinação da *identidade* de uma estrutura daquele papel que regula a sua *formação*. Na ausência de função de todos os genes *Hox*, os segmentos são formados, mas todos apresentam a mesma identidade; os membros também podem se formar, mas apresentam a identidade de antenas; e, de modo semelhante, as asas podem se formar, mas apresentam a identidade de asas dianteiras. Outros genes controlam a formação de segmentos, membros e asas e serão descritos posteriormente. Primeiramente, precisamos compreender como os genes *Hox* exercem seus efeitos dramáticos sobre o desenvolvimento da mosca.

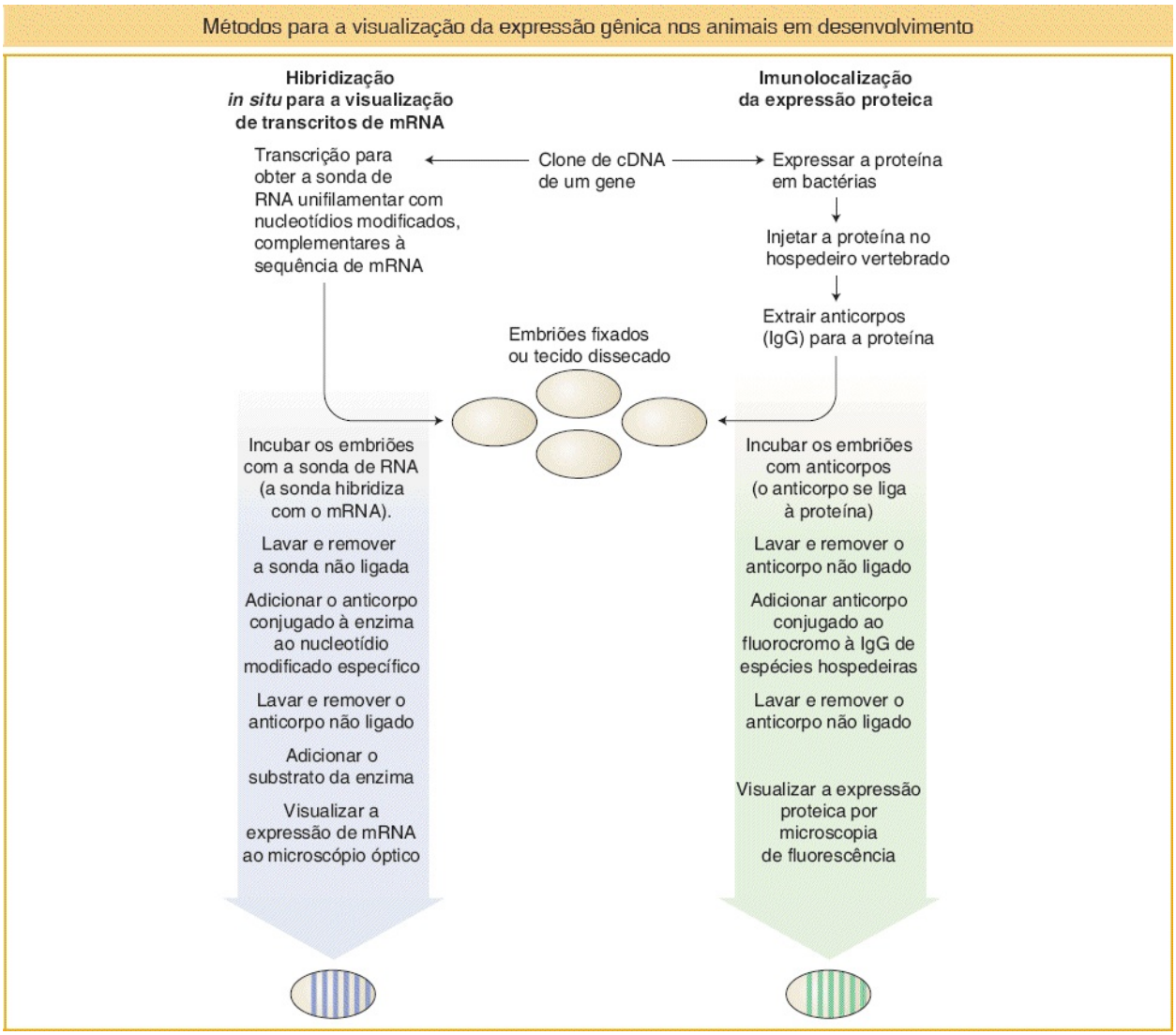


FIGURA 13.4 As duas tecnologias principais para visualizar onde um gene é transcrito ou onde a proteína que ele codifica é expressa, são (*esquerda*) a hibridização *in situ* de sonda de RNA complementar ao mRNA e (*direita*) a imunolocalização da expressão proteica. Estão resumidos os procedimentos em relação a cada método. Os padrões de expressão podem ser visualizados como o produto de uma reação enzimática ou de um substrato cromogênico ou com compostos marcados com fluorescência.

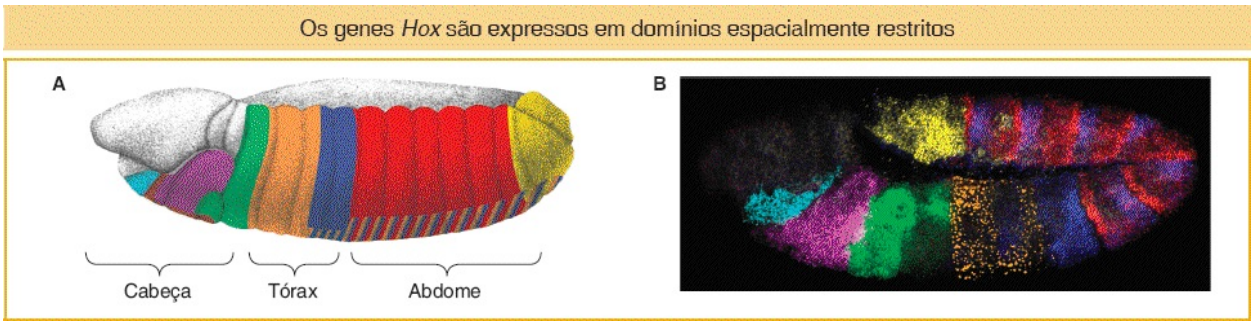


FIGURA 13.5 Expressão dos genes *Hox* no embrião de *Drosophila*. **A.** Representação esquemática do embrião de *Drosophila*, demonstrando as regiões nas quais oito genes *Hox* individuais são expressos. **B.** Imagem real da expressão de sete genes *Hox* visualizados por meio de hibridização *in situ*. As cores indicam a expressão de *labial* (*turquesa*), *Deformed* (*lavanda*), *Sex combs reduced* (*verde*), *Antennapedia* (*laranja*), *Ultrabithorax* (*azul-escuro*), *Abdominal-A* (*vermelho*) e *Abdominal-B* (*amarelo*). O embrião está dobrado, de modo que a extremidade posterior (*amarela*) aparece próximo ao centro da parte superior. (*B.* Dave Kosman, Ethan Bier e Bill McGinnis.)

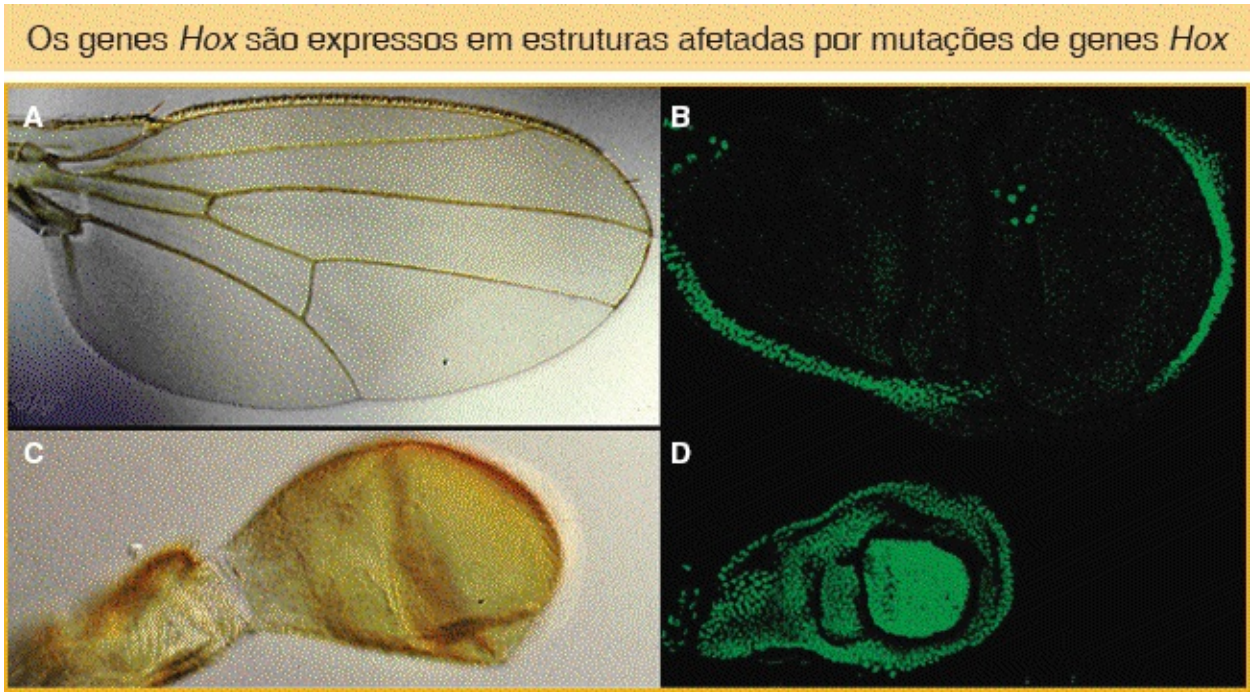


FIGURA 13.6 Um exemplo de expressão de gene *Hox*. **A.** A asa adulta de *D. melanogaster*. **B.** A proteína Ubx não é expressa em células do disco imaginal em desenvolvimento que formará as asas dianteiras. As células enriquecidas com proteínas Hox estão coradas em verde; nesta imagem, as células coradas em verde são células que *não* formam a asa. **C.** A asa traseira adulta (haltere). **D.** A proteína Ubx é expressa em níveis altos em todas as células do disco imaginal das asas traseiras em desenvolvimento. (Scott Weatherbee.)

Homeoboxe

Tendo em vista que os genes *Hox* apresentam grandes efeitos nas identidades de segmentos inteiros e outras estruturas corporais, a natureza e a função das proteínas que eles codificam são de interesse em especial. Edward Lewis, um pioneiro no estudo de genes homeóticos, observou inicialmente que o agrupamento de genes do complexo *Bithorax* sugeria que os múltiplos *loci* haviam surgido por duplicação em *tandem* de um gene ancestral. Essa ideia levou

os pesquisadores a procurar por semelhanças nas sequências de DNA dos genes *Hox*. Eles observaram que todos os oito genes *Hox* dos dois complexos eram semelhantes o suficiente para hibridizar entre si. Observou-se que essa hibridização ocorre em virtude de uma curta região da sequência em cada gene, de 180 pb de comprimento. Esse trecho de similaridade na sequência de DNA, em virtude da sua presença em genes homeóticos, foi apelidado **homeoboxe**. O homeoboxe codifica um domínio proteico, o **homeodomínio**, que contém 60 aminoácidos. A sequência de aminoácidos do homeodomínio é muito semelhante entre as proteínas Hox.

Embora a descoberta de um motivo proteico em comum em cada uma das proteínas Hox tenha sido muito empolgante, análise adicional da estrutura do homeodomínio revelou que ela forma um *motif* hélice–giro–hélice — a estrutura comum ao repressor Lac, ao repressor λ, Cro e às proteínas reguladoras α2 e a1 dos *loci* de tipos reprodutivos de leveduras! Essa semelhança sugeriu imediatamente (e foi subsequentemente corroborada) que as proteínas Hox são proteínas sequência-específicas de ligação ao DNA e que elas exercem seus efeitos por meio do controle da expressão de genes nos segmentos e apêndices em desenvolvimento. Portanto, os produtos desses genes extraordinários atuam por meio de princípios que já são familiares dos Capítulos 11 e 12 — por meio da ligação aos elementos reguladores de outros genes para ativar ou reprimir a sua expressão. Veremos que isso também é verdadeiro para muitos outros genes *toolkit*: uma fração significativa desses genes codifica fatores de transcrição que controlam a expressão de outros genes.

> CONCEITO-CHAVE Muitos genes *toolkit* codificam fatores de transcrição que regulam a expressão de outros genes.

Examinaremos como as proteínas Hox e outras proteínas toolkit orquestram a expressão gênica no desenvolvimento um pouco mais posteriormente. Primeiramente, existe mais uma grande descoberta a ser descrita, que revelou que o que aprendemos a partir dos genes *Hox* da mosca apresenta implicações muito

gerais para o reino animal.

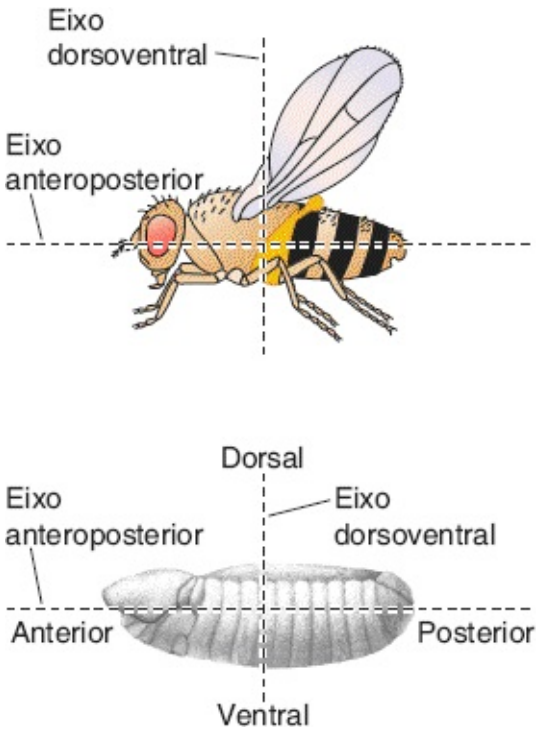
Grupos de genes Hox controlam o desenvolvimento na maior parte dos animais

Quando o homeoboxe foi descoberto nos genes *Hox* de moscas, levantou-se a dúvida sobre se essa característica era alguma peculiaridade desses genes bizarros de moscas ou se estava mais amplamente distribuída, em outros insetos ou animais segmentados, por exemplo. Para abordar essa possibilidade, pesquisadores procuraram por homeoboxes nos genomas de outros insetos, bem como de minhocas, rãs, vacas e até mesmo seres humanos. Eles encontraram muitos homeoboxes nos genomas de cada um desses animais.

As similaridades nas sequências homeoboxe de diferentes espécies eram impressionantes. Ao longo dos 60 aminoácidos do homeodomínio, algumas proteínas Hox de camundongos e rãs eram idênticas às sequências das moscas em até 59 das 60 posições. À luz das vastas distâncias evolutivas entre esses animais, com mais de 500 milhões de anos desde o seu último ancestral comum, a extensão da similaridade das sequências indica uma pressão muito forte para manter a sequência do homeodomínio.

A existência de genes *Hox* com homeoboxes em todo o reino animal era totalmente inesperada. Por que diferentes tipos de animais possuiriam os mesmos genes reguladores não era óbvio, motivo pelo qual os biólogos foram ainda mais surpreendidos pelos resultados quando a organização e a expressão de genes *Hox* foram examinadas em outros animais. Em vertebrados, como o camundongo de laboratório, os genes *Hox* também estão agrupados em quatro grandes complexos gênicos em quatro cromossomos diferentes. Cada grupo contém de 9 a 11 genes *Hox*, um total de 39 genes *Hox*. Além disso, a ordem dos genes nos complexos *Hox* de camundongos é paralela à ordem de seus correspondentes mais correlatos nos complexos *Hox* de moscas, bem como em cada um dos outros grupos *Hox* de camundongos. Essa correspondência indica que os complexos *Hox* de insetos e vertebrados estão relacionados e que existia algum tipo de complexo *Hox* em seu ancestral comum distante. Os quatro complexos *Hox* em camundongos surgiram por meio de duplicações de complexos *Hox* inteiros (talvez de cromossomos

inteiros) em ancestrais dos vertebrados.



A relação entre os eixos corporais adulto e embrionário

Por que tais animais diferentes apresentariam esses conjuntos de genes em comum? Seus ancestrais comuns indicam que os genes *Hox* desempenham algum papel fundamental no desenvolvimento da maior parte dos animais. Esse papel está aparente a partir de análises sobre como os genes *Hox* são expressos em diferentes animais. Em embriões de vertebrados, os genes *Hox* adjacentes também são expressos em domínios adjacentes ou parcialmente sobrepostos ao longo do eixo corporal anteroposterior. Além disso, a ordem dos genes *Hox* nos complexos corresponde à ordem da cabeça à cauda das regiões corporais nas quais os genes são expressos.

Os padrões de expressão dos genes *Hox* de vertebrados sugeriam que eles também especificam a identidade das regiões corporais e análises subsequentes de mutantes dos genes *Hox* corroboraram essa sugestão. Por exemplo, mutações nos genes *Hoxa11* e *Hoxd11* causam a transformação homeótica das vértebras sacrais em vértebras lombares (Figura 13.7). Portanto, assim como na mosca,

perda ou ganho de função dos genes *Hox* em vertebrados causa a transformação da identidade de estruturas serialmente repetidas. Os referidos resultados foram obtidos em diversas classes, incluindo mamíferos, aves, anfíbios e peixes. Além disso, foi demonstrado que grupos de genes *Hox* regulam a padronização de outros insetos e estão dispostos em regiões ao longo do eixo anteroposterior em anelídeos, moluscos, nematódeos, diversos artrópodes, cordatos primitivos, platelmintos e outros animais. Portanto, apesar das enormes diferenças na anatomia, a apresentação de um ou mais grupos de genes *Hox* que estão dispostos em regiões ao longo do eixo corporal principal é uma característica comum e fundamental de no mínimo todos os animais bilaterais. De fato, as lições surpreendentes dos genes *Hox* pressagiaram o que se descobriu ser uma tendência geral entre os genes *toolkit*; ou seja, a maior parte dos genes *toolkit* é comum a diferentes animais.

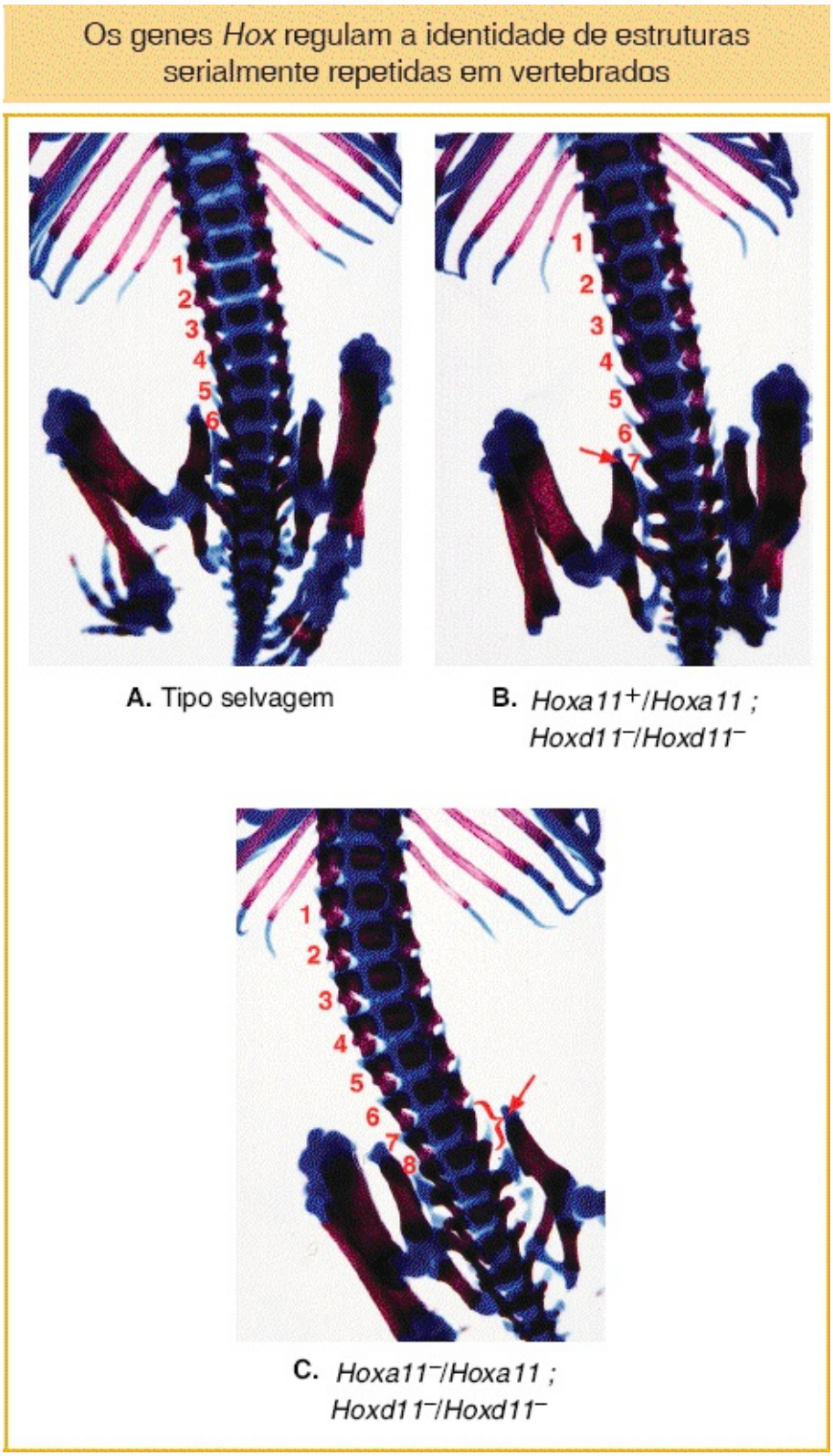


FIGURA 13.7 As morfologias das diferentes regiões da coluna vertebral são reguladas pelos genes *Hox*. **A.** Em camundongos, seis vértebras lombares são formadas logo anteriormente às vértebras sacrais (*números em*

vermelho). **B.** Em camundongos com ausência da função de ação posterior do gene *Hoxd11* e que apresentam uma cópia funcional do gene *Hoxa11*, sete vértebras lombares se formam e uma vértebra sacral é perdida. **C.** Em camundongos com ausência da função de ambos *Hoxa11* e *Hoxd11*, oito vértebras lombares se formam e duas vértebras sacrais são perdidas. (As fotografias são cortesia da Dra. Anne Boulet, HHMI, University of Utah; de S. B. Carroll, J. K. Grenier e S. D. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*, 2nd ed. Blackwell, 2005.)

> CONCEITO-CHAVE Apesar de grandes diferenças na anatomia, muitos genes *toolkit* são comuns a uma ampla variedade de diferentes filos de animais.

Agora, façamos um inventário do restante *toolkit* para verificar quais outros princípios gerais surgem.

13.3 Definição da toolkit inteira

Os genes *Hox* talvez sejam os membros mais bem-conhecidos da *toolkit*, mas são apenas uma pequena família em um grupo muito maior de genes necessários para o desenvolvimento dos números, formas, tamanhos e tipos adequados de partes do corpo. Pouco se sabia a respeito do restante da *toolkit* até o final da década de 1970 e o início da década de 1980, quando Christine Nüsslein-Volhard e Eric Wieschaus, que trabalhavam no Max Planck Institute em Tübingen, Alemanha, encontraram os genes necessários para a formação da organização segmentar do embrião e da larva de *Drosophila*.

Até os seus esforços, a maior parte dos trabalhos sobre o desenvolvimento da mosca enfocava fenótipos de adultos viáveis, não o embrião. Nüsslein-Volhard e Wieschaus perceberam que os tipos de genes que eles estavam procurando provavelmente eram letais para os embriões ou as larvas em mutantes homozigotos. Assim, eles inventaram um esquema para procurar os genes que eram necessários no **zigoto** (o produto da fertilização; Figura 13.8, *parte inferior*). Eles também desenvolveram triagens para identificar os genes com produtos que atuam no zigoto antes que o genoma zigótico esteja ativo e que são

necessários para a adequada padronização do embrião. Os genes com produtos fornecidos pela fêmea ao zigoto são denominados **genes de efeito materno**. Os fenótipos mutantes de genes de efeito materno estrito dependem apenas do genótipo da fêmea (Figura 13.8, parte superior).

Nessas triagens, foram identificados genes que eram necessários para produzir o número e o padrão adequados dos segmentos larvais, para produzir as suas três camadas teciduais (ectoderma, mesoderma e endoderma) e para padronizar os detalhes finos da anatomia animal. O poder das triagens genéticas era a sua natureza sistemática. Ao saturar cada um dos cromossomos da mosca (com exceção do pequeno quarto cromossomo) com mutações induzidas quimicamente, os pesquisadores conseguiram identificar a maior parte dos genes que eram necessários para a construção da mosca. Por seus esforços pioneiros, Nüsslein-Volhard, Wieschaus e Lewis compartilharam o Prêmio Nobel de Fisiologia ou Medicina de 1995.

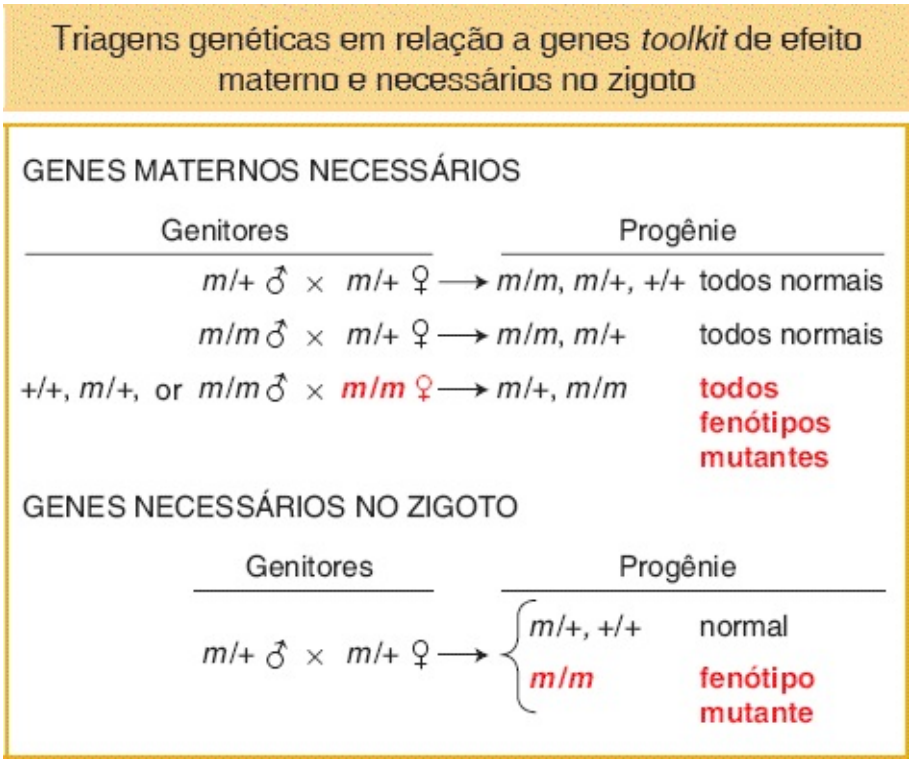


FIGURA 13.8 Triagens genéticas identificam se um produto gênico atua no ovócito ou no zigoto. Os fenótipos da progênie dependem (*parte superior*) do genótipo materno em relação aos genes de efeito materno ou (*parte inferior*) do genótipo da progênie (zigótico) em relação aos genes zigoticamente necessários (*m* =

Mutante; + = Tipo selvagem).

As características mais surpreendentes e informativas dos mutantes recentemente identificados eram que eles demonstravam defeitos dramáticos, porém bem-definidos, na organização ou na padronização do embrião. Ou seja, a larva morta não era uma carcaça amorfa, mas exibia defeitos de padronização específicos, com frequência surpreendentes. O corpo da larva de *Drosophila* apresenta diversas características cujo número, posição ou padrão podem atuar como pontos de referência para diagnosticar ou classificar as anormalidades em animais mutantes. Cada *locus*, portanto, pode ser classificado de acordo com o eixo corporal afetado e o padrão de defeitos causados pela mutação. Cruzamentos genéticos revelam se o *locus* está ativo no ovócito materno ou no zigoto. Cada classe de genes parece representar diferentes etapas no refinamento progressivo do plano corporal embrionário — desde aqueles que afetam grandes regiões do embrião, até aqueles com áreas de influência mais limitadas.

Em relação a qualquer gene *toolkit*, três informações são chaves para a compreensão da função gênica: (1) o fenótipo mutante, (2) o padrão de expressão gênica e (3) a natureza do produto gênico. Estudos extensivos de dúzias de genes levaram a um quadro razoavelmente detalhado de como cada eixo corporal é estabelecido e subdividido em segmentos ou camadas germinativas.

Eixos anteroposterior e dorsoventral

São necessárias algumas dúzias de genes para a adequada organização do eixo corporal anteroposterior do embrião da mosca. Os genes estão agrupados em cinco classes com base na sua região de influência no padrão embrionário.

- A primeira classe estabelece o eixo anteroposterior e é composta pelos genes de efeito materno. Um membro-chave dessa classe é o gene *Bicoid*. Embriões de mães mutantes *Bicoid* apresentam ausência da região anterior do embrião (Figura 13.9), o que nos informa que o gene é necessário para o desenvolvimento daquela região.

As próximas três classes são genes zigoticamente ativos necessários para o

desenvolvimento dos segmentos do embrião.

- A segunda classe contém os **genes *gap***. Cada um desses genes afeta a formação de um bloco contíguo de segmentos; mutações nos genes *gap* levam a grandes intervalos na segmentação (Figura 13.10, esquerda)
- A terceira classe abrange os **genes de regra dos pares**, que atuam na periodicidade de segmentos duplos. Os mutantes de regra dos pares apresentam ausência de uma parte de cada par de segmentos, mas diferentes genes de regra dos pares afetam diferentes partes de cada segmento duplo. Por exemplo, o gene *even-skipped* afeta um conjunto de limites segmentares e o gene *odd-skipped* afeta o conjunto complementar de limites (Figura 13.10, *parte intermediária*)

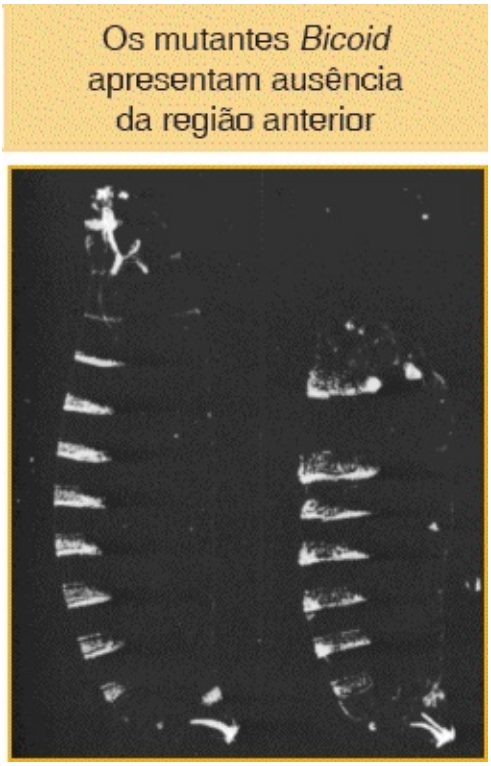


FIGURA 13.9 O gene de efeito materno *Bicoid* (*bcd*) afeta a parte anterior da larva em desenvolvimento. Estas fotomicrografias são de larvas de *Drosophila* que foram preparadas para demonstrar seus exoesqueletos rígidos. Estruturas densas, tais como as bandas denticuladas segmentares, aparecem em branco. (*Esquerda*) Uma larva normal. (*Direita*) Larva de uma fêmea mutante *bcd* homozigota. A cabeça e as estruturas torácicas anteriores estão ausentes. (De C. H. Nüsslein-Volhard, G. Frohnhöfer e R. Lehmann, "Determination of Anteroposterior Polarity in Drosophila", *Science* 238, 1987, 1678. Reproduzida, com autorização, da AAAS.)

- A quarta classe é composta por **genes de polaridade do segmento**, que afetam a padronização dentro de cada segmento. Mutantes dessa classe demonstram defeitos na polaridade e no número de segmentos (Figura 13.10, *direita*).

A quinta classe de genes determina o destino de cada segmento.

- A quinta classe inclui os genes *Hox* já discutidos; os mutantes *Hox* não afetam o número de segmentos, mas alteram o aspecto de um ou mais segmentos.

Expressão dos genes *toolkit*

Para compreender a relação entre genes e o fenótipo mutante, precisamos saber a ocasião e a localização dos padrões de expressão dos genes e a natureza molecular dos produtos gênicos. Os padrões de expressão dos genes *toolkit* correspondem de modo nítido aos seus fenótipos, uma vez que com frequência estão precisamente correlacionados com as partes do corpo em desenvolvimento alteradas nos mutantes. Cada gene é expresso em uma região que pode ser mapeada para coordenadas específicas ao longo de cada eixo do embrião. Por exemplo, a proteína Bicoid de efeito materno é expressa em um padrão graduado que emana do polo anterior do embrião inicial, a seção do embrião ausente nos mutantes (Figura 13.11 A). De modo semelhante, as proteínas gap são expressas em blocos de células que correspondem às futuras posições dos segmentos que estão ausentes nos respectivos mutantes de genes *gap* (Figura 13.11 B). As proteínas de regra dos pares são expressas em padrões listrados surpreendentes: uma listra transversal é expressa a cada dois segmentos, em um total de sete listras, que abrangem os 14 segmentos corporais futuros (a posição e a periodicidade das listras correspondem à periodicidade dos defeitos em larvas mutantes), conforme demonstrado na Figura 13.11 C. Muitos genes de polaridade do segmento são expressos em listras de células dentro de cada segmento, 14 listras que no total correspondem aos 14 segmentos corporais (Figura 13.11 D). Observe que os domínios da expressão gênica se tornam progressivamente mais refinados na medida em que o desenvolvimento prossegue: os genes são expressos primeiramente em grandes regiões (proteínas gap), em seguida em

listras com três a quatro células de largura (proteínas de regra dos pares) e em seguida em listras com uma a duas células de largura (proteínas de polaridade do segmento).

Além do que aprendemos a partir dos padrões espaciais da expressão de genes *toolkit* a ordem de expressão desses genes ao longo do tempo é lógica. A proteína Bicoid de efeito materno aparece antes das proteínas gap zigóticas, que são expressas antes do aparecimento dos padrões de sete listras das proteínas de regra dos pares, que por sua vez precedem os padrões de 14 listras das proteínas de polaridade do segmento. A ordem da expressão gênica e o progressivo refinamento dos domínios no embrião revelam que a produção do plano corporal é um processo passo a passo, com subdivisões importantes do corpo primeiramente delineadas e em seguida refinadas até que um padrão fino seja estabelecido. A ordem de ação gênica sugere adicionalmente que a expressão de um conjunto de genes pode regular a expressão do conjunto de genes seguintes.

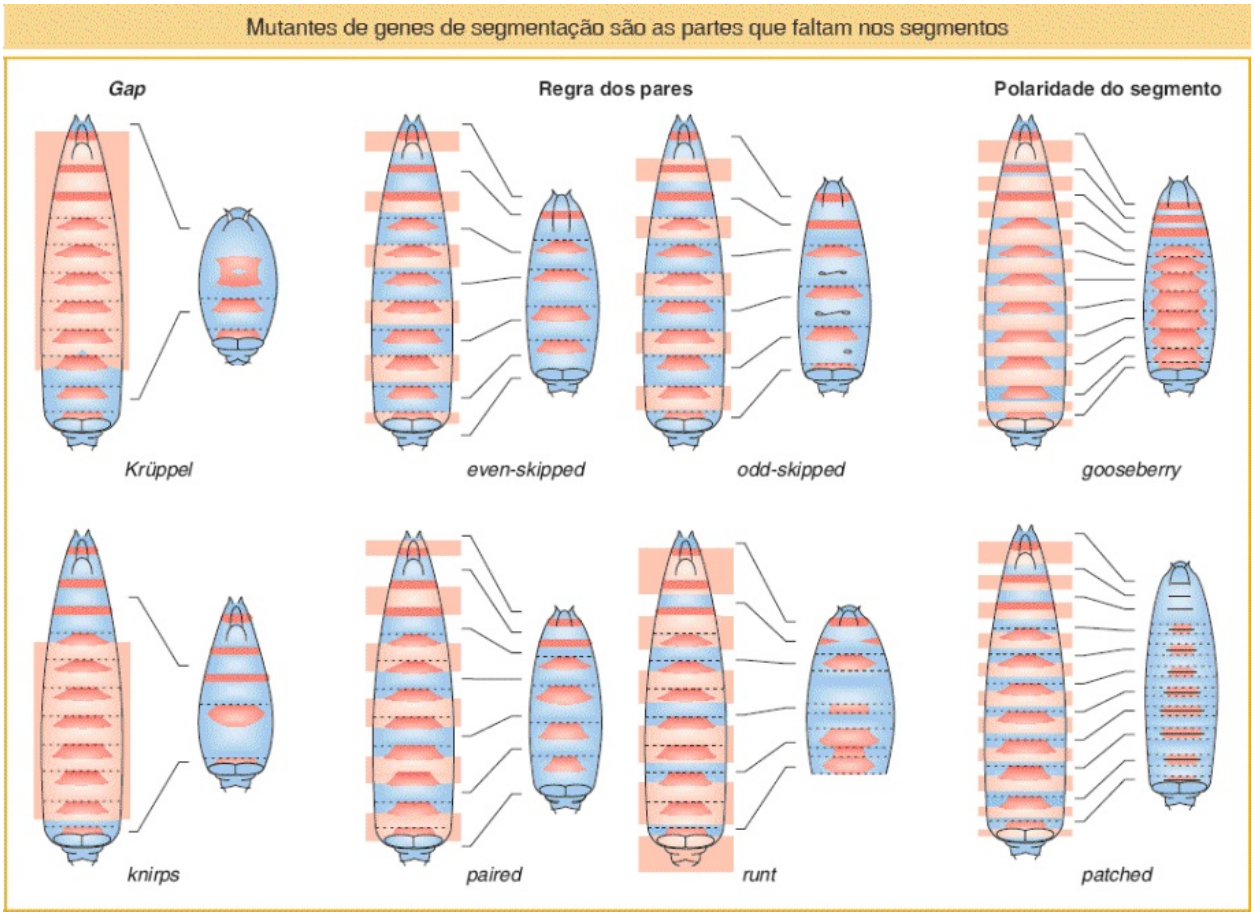


FIGURA 13.10 Classes de mutantes de genes de segmentação de *Drosophila*. Esses diagramas ilustram mutantes gap, regra dos pares e polaridade do segmento representativos. Os *trapézios vermelhos* são as faixas densas do exoesqueleto observado na Figura 13.9. O limite de cada segmento está indicado por uma *linha pontilhada*. O diagrama à esquerda de cada par ilustra uma larva do tipo selvagem e o diagrama à direita ilustra o padrão formado em um determinado mutante. As regiões *laranja-claras sombreadas* nos diagramas do tipo selvagem ilustram os domínios da larva que estão ausentes ou afetados no mutante.

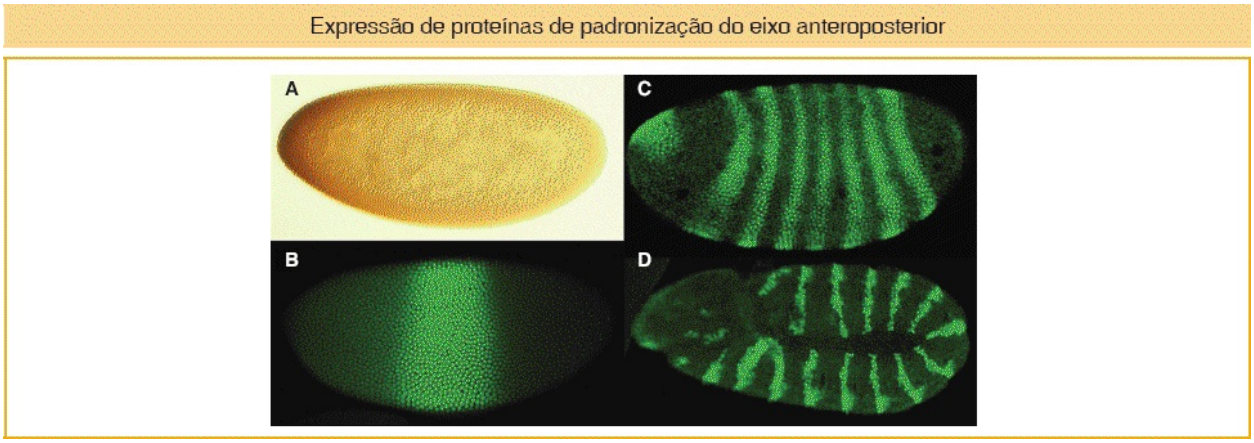


FIGURA 13.11 Os padrões de expressão de genes *toolkit* correspondem a fenótipos mutantes. Embriões de *Drosophila* foram corados com anticorpos contra: **A.** a proteína Bicoid de origem materna; **B.** a proteína gap Krüppel; **C.** a proteína de regra dos pares Hairy; **D.** a proteína de polaridade do segmento Engrailed. Os embriões foram visualizados por meio de métodos imunoenzimáticos (*a coloração é marrom*) (**A**) ou com imunofluorescência (a coloração é *verde*) (**B** a **D**). Cada proteína está localizada nos núcleos em regiões do embrião que são afetadas por mutações nos respectivos genes. (As fotomicrografias são cortesia de: *A.* Ruth Lehmann; *B* a *D.* James Langeland.)

Uma indicação de que essa progressão de fato é o que ocorre advém da análise dos efeitos das mutações em genes *toolkit* na expressão de outros genes *toolkit*. Por exemplo, em embriões de mães mutantes *Bicoid*, a expressão de diversos genes *gap* está alterada, bem como aquela dos genes de regra dos pares e de polaridade do segmento. Esse achado sugere que a proteína Bicoid de algum modo (direta ou indiretamente) influencia a regulação dos genes *gap*.

Outra indicação de que a expressão de um conjunto de genes pode regular a expressão do conjunto de genes que se sucede advém do exame dos produtos proteicos. A inspeção da sequência da proteína Bicoid revela que ela contém um homeodomínio, relacionado com o das proteínas Hox, porém distinto dele. Portanto, a Bicoid apresenta as propriedades de um fator de transcrição de

ligação ao DNA. Cada gene *gap* também codifica um fator de transcrição, assim como cada gene de regra dos pares, diversos genes de polaridade do segmento e, conforme descrito anteriormente, todos os genes *Hox*. Esses fatores de transcrição incluem representantes das famílias mais conhecidas de proteínas sequência-específicas de ligação ao DNA; assim, embora não exista restrição a respeito de a qual família elas possam pertencer, muitas proteínas toolkit de ação precoce são fatores de transcrição. Aquelas que não são fatores de transcrição tendem a ser componentes de vias de sinalização (Tabela 13.1). Essas vias medeiam processos de sinalização induzidos por ligantes entre as células e seu produto em geral leva à ativação ou à repressão gênica. Portanto, a maior parte das proteínas toolkit afeta a regulação gênica, seja direta (como fatores de transcrição) ou indiretamente (como componentes de vias de sinalização). A maior parte das vias de sinalização opera por meio de lógicas semelhantes, mas apresenta diferentes componentes proteicos e mecanismos de transdução de sinal.

> CONCEITO-CHAVE A maior parte das proteínas toolkit são fatores de transcrição ou componentes de vias de transdução de sinal mediadas por ligantes.

Os mesmos princípios se aplicam à produção do eixo corporal dorsoventral, assim como se aplicam ao eixo anteroposterior. O eixo dorsoventral também é subdividido em regiões. Diversos genes de efeito materno, tais como *dorsal*, são necessários para estabelecer essas regiões em posições distintas do lado dorsal (superior) em relação ao ventral (inferior) do embrião. Os mutantes *dorsais* são "dorsalizados" e não apresentam estruturas ventrais (tais como o mesoderma e o sistema nervoso). Um punhado de genes que são ativados no zigoto também são necessários para a subdivisão do eixo dorsovenral.

O produto do gene de efeito materno *dorsal* é um fator de transcrição — a proteína Dorsal. Essa proteína é expressa em um gradiente ao longo do eixo dorsoventral, com seu mais alto nível de acúmulo em células ventrais (Figura 13.12 A). O gradiente estabelece sub-regiões de diferentes concentrações da

proteína Dorsal. Em cada sub-região, é expresso um conjunto diferente de genes zigóticos que contribuem para a padronização dorsoventral. Os conjuntos de genes zigóticos expressos definem regiões que dão origem a camadas teciduais em particular, tais como o mesoderma e o neuroectoderma (a parte do ectoderma que dá origem ao sistema nervoso ventral), conforme demonstrado na Figura 13.12 B.

Tabela 13.1 Exemplos de genes do eixo A-P de *Drosophila* que contribuem para a formação de padrões.

Símbolo do gene	**Nome do gene**	**Função proteica**	**Papel(éis) no desenvolvimento inicial**
hb-z	*hunchback-zygotic*	Fator de transcrição — proteína zinc-finger	Gene *gap*
Kr	*Krüppel*	Fator de transcrição — proteína zinc-finger	Gene *gap*
kni	*knirps*	Fator de transcrição — proteína do tipo receptor de esteroide	Gene *gap*
eve	*even-skipped*	Fator de transcrição — proteína de homeodomínio	Gene de regra dos pares
ftz	*fushi tarazu*	Fator de transcrição — proteína de	Gene de regra dos

		homeodomínio	pares
opa	*odd-paired*	Fator de transcrição — proteína zinc-finger	Gene de regra dos pares
prd	*paired*	Fator de transcrição — proteína PHOX	Gene de regra dos pares
en	*engrailed*	Fator de transcrição — proteína de homeodomínio	Gene de polaridade do segmento
wg	*wingless*	Proteína de sinalização WG	Gene de polaridade do segmento
hh	*hedgehog*	Proteína de sinalização HH	Gene de polaridade do segmento
ptc	*patched*	Proteína transmembrana	Gene de polaridade do segmento
lab	*labial*	Fator de transcrição — proteína de homeodomínio	Gene de identidade do segmento
Dfd	*Deformed*	Fator de transcrição — proteína de homeodomínio	Gene de identidade do segmento

Antp	*Antennapedia*	Fator de transcrição — proteína de homeodomínio	Gene de identidade do segmento
Ubx	*Ultrabithorax*	Fator de transcrição — proteína de homeodomínio	Gene de identidade do segmento

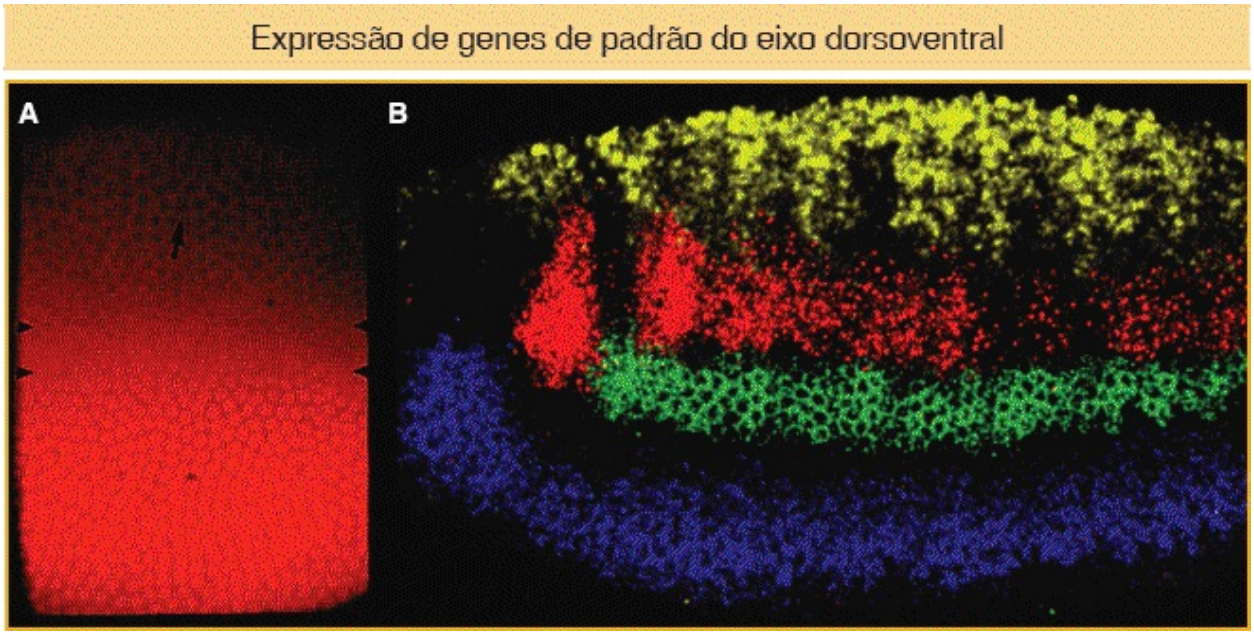


FIGURA 13.12 Os domínios de expressão dos genes de padrão do eixo dorsoventral específicos correspondem a camadas específicas de futuros tecidos. **A.** A proteína Dorsal de origem materna é expressa em um gradiente, com a mais alta concentração de Dorsal nos núcleos de células ventrais (*parte inferior da fotografia*). **B.** A expressão de quatro genes zigóticos de padrão do eixo dorsoventral revelada por meio de hibridização *in situ* com RNA. Nesta vista lateral, estão revelados os domínios dos genes *decapentaplegic* (*amarelo*), *muscle segment homeobox* (*vermelho*), *intermediate neuroblasts defective* (*verde*) e *ventral neuroblasts defective* (*azul*). (Cortesia de: *A.* Michael Levine; *B.* David Kosnan, Bill McGinnis e Ethan Bier.)

O controle genético do desenvolvimento é, então, fundamentalmente uma

questão de regulação gênica no espaço e no tempo. Como a ligação e o desligamento de genes *toolkit* constroem a forma animal? E como isso é coreografado durante o desenvolvimento? Para responder a essas questões, examinaremos as interações das proteínas com os genes *toolkit* da mosca em mais detalhes. Os mecanismos que veremos em relação ao controle da expressão dos genes *toolkit* no embrião de *Drosophila* surgiram como modelos para a regulação espacial da expressão gênica no desenvolvimento animal em geral.

13.4 Regulação espacial da expressão gênica no desenvolvimento

Vimos que os genes *toolkit* são expressos em referência a coordenadas no embrião. Mas como as coordenadas espaciais do embrião em desenvolvimento são repassadas como instruções para os genes, para ligá-los ou desligá-los em padrões precisos? Conforme descrito nos Capítulos 11 e 12, o controle fisiológico da expressão gênica em bactérias e eucariotos simples é, em última instância, regulado por proteínas sequência-específicas de ligação ao DNA que atuam em elementos reguladores de ação cis (p. ex., operadores e elementos de sequência de ativação *upstream*, ou UAS). De modo semelhante, o controle espacial da expressão gênica durante o desenvolvimento é amplamente regulado pela interação de fatores de transcrição com os elementos reguladores de ação cis. Entretanto, o controle espacial e temporal da regulação gênica no desenvolvimento de um embrião multicelular tridimensional requer a ação de mais fatores de transcrição em elementos reguladores de ação cis mais numerosos e mais complexos.

Para definir uma posição em um embrião, deve existir informação reguladora que diferencie aquela posição das regiões adjacentes. Se ilustrarmos um embrião tridimensional como um globo, deve então ser especificada a **informação posicional** que indique a longitude (localização ao longo do eixo anteroposterior), a latitude (localização ao longo do eixo dorsoventral) e a altitude ou profundidade (posição nas camadas germinativas). Ilustraremos os princípios

gerais a respeito de como as posições da expressão gênica são especificadas com três exemplos. Esses exemplos devem ser considerados apenas como algumas ilustrações do vasto número de interações regulatórias que coordenam o desenvolvimento das moscas e dos animais. O desenvolvimento é um contínuo no qual cada padrão de atividade gênica apresenta uma base causal precedente. O processo inteiro inclui dezenas de milhares de interações regulatórias e produtos.

Enfocaremos algumas conexões entre os genes em diferentes níveis de hierarquia que determinam o plano corporal segmentar básico e os *pontos nodais* nos quais genes-chave integram múltiplos estímulos reguladores e respondem produzindo expressão gênica mais simples.

Gradientes maternos e ativação gênica

A proteína Bicoid é um fator de transcrição do tipo homeodomínio que é traduzida a partir do mRNA de origem materna depositado no zigoto e localizado no polo anterior. Tendo em vista que o embrião inicial de *Drosophila* é um *sincício* com todos os núcleos em um citoplasma e ausência de quaisquer membranas celulares que possam impedir a difusão de moléculas proteicas, a proteína Bicoid consegue se difundir pelo citoplasma. Essa difusão estabelece um gradiente de concentração proteica (Figura 13.13 A): a proteína Bicoid está altamente concentrada na extremidade anterior e essa concentração diminui gradativamente na medida em que a distância daquela extremidade aumenta, até que haja muito pouca proteína Bicoid além da parte intermediária do embrião. Esse gradiente de concentração fornece informação posicional a respeito da localização ao longo do eixo anteroposterior. Uma concentração alta significa a extremidade anterior, uma concentração mais baixa significa a parte intermediária e assim por diante. Portanto, um modo de assegurar que um gene esteja ativado em apenas uma localização ao longo do eixo é relacionar a expressão gênica ao nível de concentração. Um caso em questão é o dos genes *gap*, que devem ser ativados em regiões específicas ao longo do eixo.

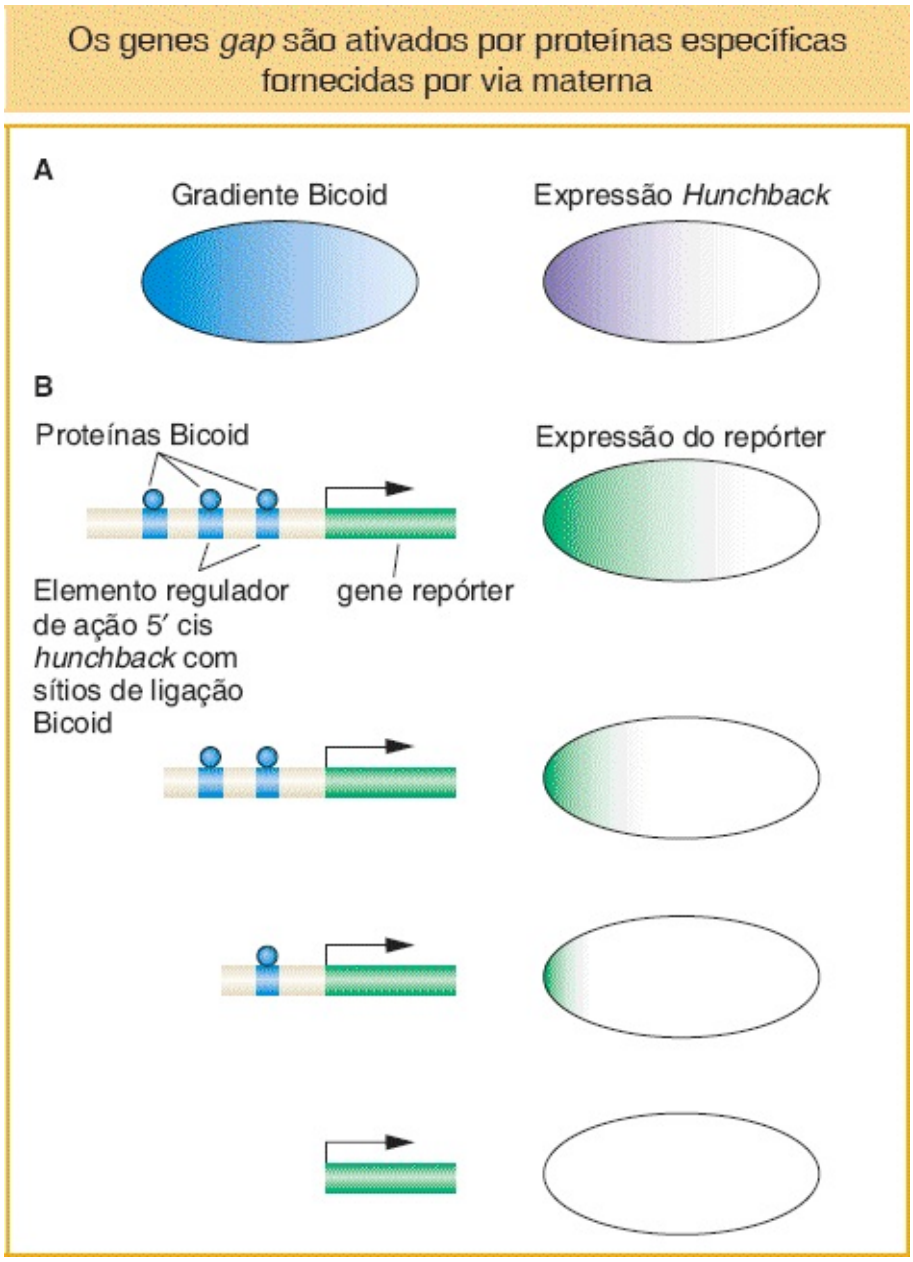


FIGURA 13.13 A proteína Bicoid ativa a expressão zigótica do gene *hunchback*. **A.** A expressão da proteína Bicoid é graduada ao longo do eixo anteroposterior. O gene *gap hunchback* é expresso na metade anterior do zigoto. **B.** A proteína Bicoid (*azul*) se liga a três sítios 5′ do gene *hunchback*. Quando esse DNA 5′ está posicionado *upstream* de um gene repórter, a *expressão* do gene repórter recapitula o padrão de expressão de *hunchback* (*parte superior à direita*). Entretanto, a deleção progressiva de um, dois ou todos os três sítios de ligação à Bicoid leva à expressão mais restrita do gene repórter ou a suprime totalmente. Estas observações demonstram que o nível e o padrão de expressão de *hunchback* são controlados por Bicoid por meio de sua ligação às sequências reguladoras do DNA *hunchback*.

Diversos genes zigóticos, incluindo os genes *gap*, são regulados por diferentes

níveis da proteína Bicoid. Por exemplo, o gene *hunchback* é um gene *gap* ativado no zigoto na metade anterior do embrião. Essa ativação ocorre por meio da ligação direta da proteína Bicoid a três sítios 5′ do promotor do gene *hunchback*. Bicoid se liga a esses sítios *de modo cooperativo*; ou seja, a ligação de uma molécula de proteína Bicoid a um sítio facilita a ligação de outras moléculas Bicoid a sítios próximos.

Experimentos *in vivo* podem demonstrar que a ativação de *hunchback* depende do gradiente de concentração. Esses testes requerem a ligação de sequências reguladoras do gene a um gene repórter (um gene codificador de enzima, tal como o gene *LacZ* ou a proteína fluorescente verde de água-viva), introduzindo o construto de DNA na linhagem germinativa da mosca e monitorando a expressão do repórter na progênie de embriões das moscas transgênicas (uma visão geral do método está demonstrada na Figura 13.14). As sequências 5′ do tipo selvagem do gene *hunchback* são suficientes para direcionar a expressão do repórter na metade anterior do embrião. De importância, deleções de sítios de ligação da Bicoid nesse elemento regulador de ação cis reduzem ou eliminam a expressão do repórter (ver Figura 13.13 B). Mais de um sítio da Bicoid deve estar ocupado para gerar uma delimitação nítida da expressão do repórter, que indica que uma concentração limiar da proteína Bicoid é necessária para ocupar múltiplos sítios antes da ativação da expressão gênica. Um gene *gap* com menos sítios de ligação não será ativado em locais com concentrações mais baixas da proteína Bicoid.

Cada gene *gap* contém elementos reguladores de ação cis com diferentes arranjos de sítios de ligação e esses sítios de ligação podem apresentar afinidades diferentes pela proteína Bicoid. Consequentemente, cada gene *gap* é expresso em um domínio distinto único no embrião, em resposta a níveis diferentes de Bicoid e outros gradientes de fatores de transcrição. Um tema semelhante é encontrado na padronização do eixo dorsoventral: os elementos reguladores de ação cis contêm diferentes números e arranjos de sítios de ligação para Dorsal e outros fatores de transcrição dorsoventral. Consequentemente, os genes são ativados em domínios discretos ao longo do eixo dorsoventral.

> CONCEITO-CHAVE A resposta dependente de concentração dos genes a estímulos graduados é uma característica crucial da regulação gênica no embrião inicial de *Drosophila*. Os elementos reguladores de ação cis que controlam respostas distintas contêm diferentes números e arranjos de sítios de ligação de fatores de transcrição.

Desenho das listras | Integração dos estímulos das proteínas gap

A expressão de cada gene de regra dos pares em sete listras é o primeiro sinal da organização periódica do embrião e do futuro animal. Como os referidos padrões periódicos são gerados a partir de uma informação aperiódica anterior? Antes da análise molecular da regulação dos genes de regra dos pares, foram apresentados diversos modelos para explicar a formação das listras. Cada uma dessas ideias visualizava todas as sete listras como produtos idênticos em resposta a estímulos idênticos. Entretanto, o modo real pelo qual os padrões de alguns genes de regra dos pares importantes são codificados e gerados é uma listra por vez. A solução para o mistério da geração das listras destaca um dos conceitos mais importantes a respeito do controle espacial da regulação gênica nos animais em desenvolvimento; a saber, os elementos reguladores de ação cis distintos de genes individuais são controlados independentemente.

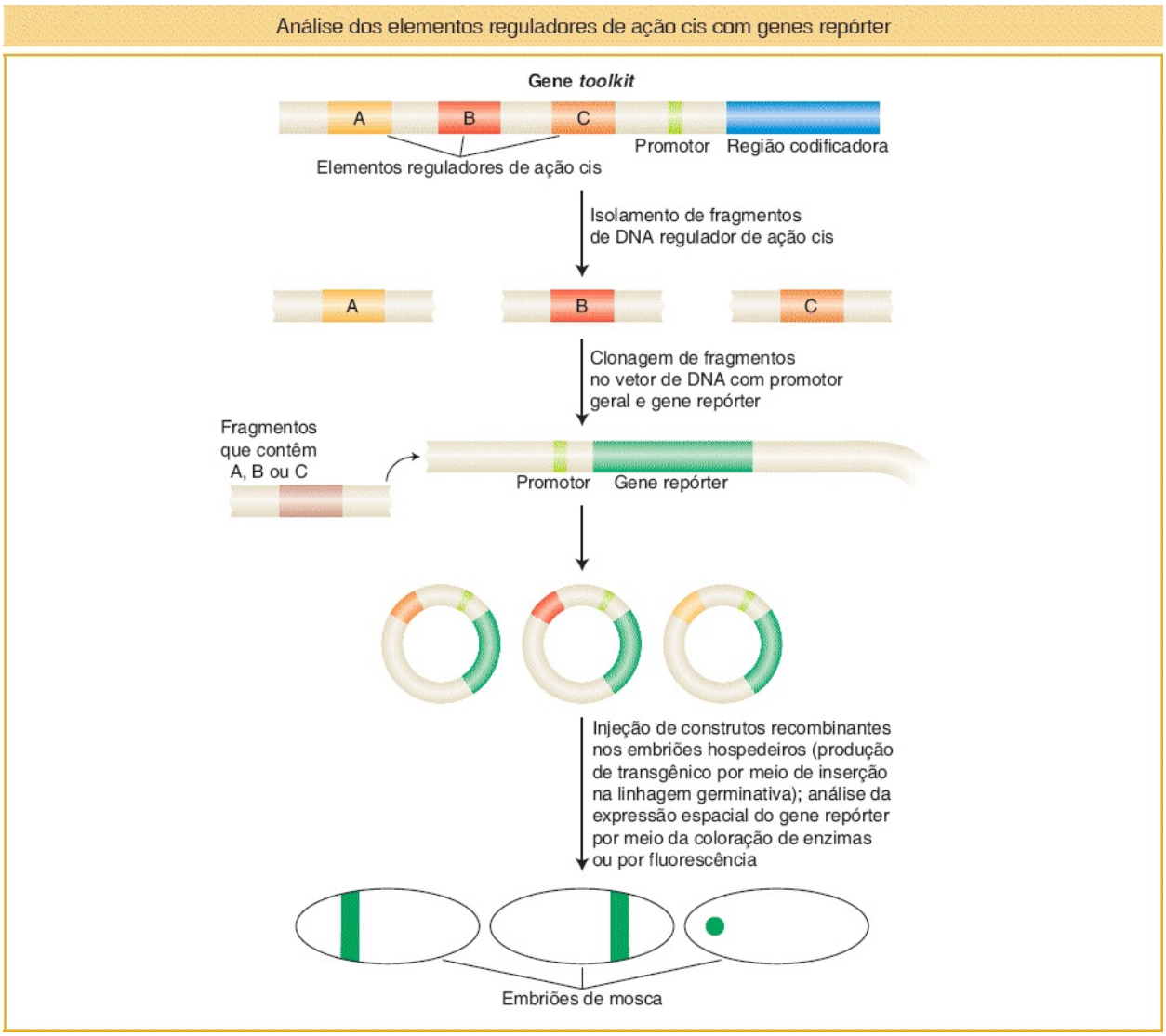


FIGURA 13.14 Os *loci toolkit* (tais como *hunchback*, conforme descrito no texto) com frequência contêm múltiplos elementos reguladores de ação cis independentes que controlam a expressão gênica em diferentes locais, em diferentes ocasiões durante o desenvolvimento ou em ambos (p. ex., A, B, C, aqui). Esses elementos são identificados por sua capacidade, quando posicionados em cis em um gene repórter e inseridos de volta em um genoma hospedeiro, de controlar o padrão, a ocasião ou o nível ou todos os três, da expressão do gene repórter. Nesse exemplo, cada elemento direciona um padrão diferente de expressão gênica em um embrião de mosca. A maior parte dos genes repórter codifica enzimas ou proteínas fluorescentes que podem ser facilmente visualizadas.

A descoberta-chave foi de que cada uma das sete listras que compõem os padrões de expressão dos genes de regra dos pares *even-skipped* e *hairy* é controlada independentemente. Considere a segunda listra expressa pelo gene *even-skipped* (Figura 13.15 A). Essa listra está localizada dentro da região ampla

de expressão de *hunchback* e nas bordas das regiões de expressão de duas outras proteínas gap, Giant e Krüppel (Figura 13.15 B). Portanto, dentro da área da futura listra, haverá grandes quantidades de proteínas Hunchback e pequenas quantidades de proteínas Giant e Krüppel. Também haverá uma determinada concentração da proteína Bicoid de efeito materno. Nenhuma outra listra no embrião conterá essas proteínas nessas proporções. A formação da listra 2 é controlada por meio de um elemento regulador de ação cis específico, um acentuador, que contém um número de sítios de ligação para essas quatro proteínas (Figura 13.15 C). A análise detalhada do elemento regulador de ação cis *eve* da listra 2 revelou que essa listra "simples" é controlada pela ligação desses quatro fatores de transcrição aperiodicamente distribuídos, incluindo uma proteína materna e três proteínas gap.

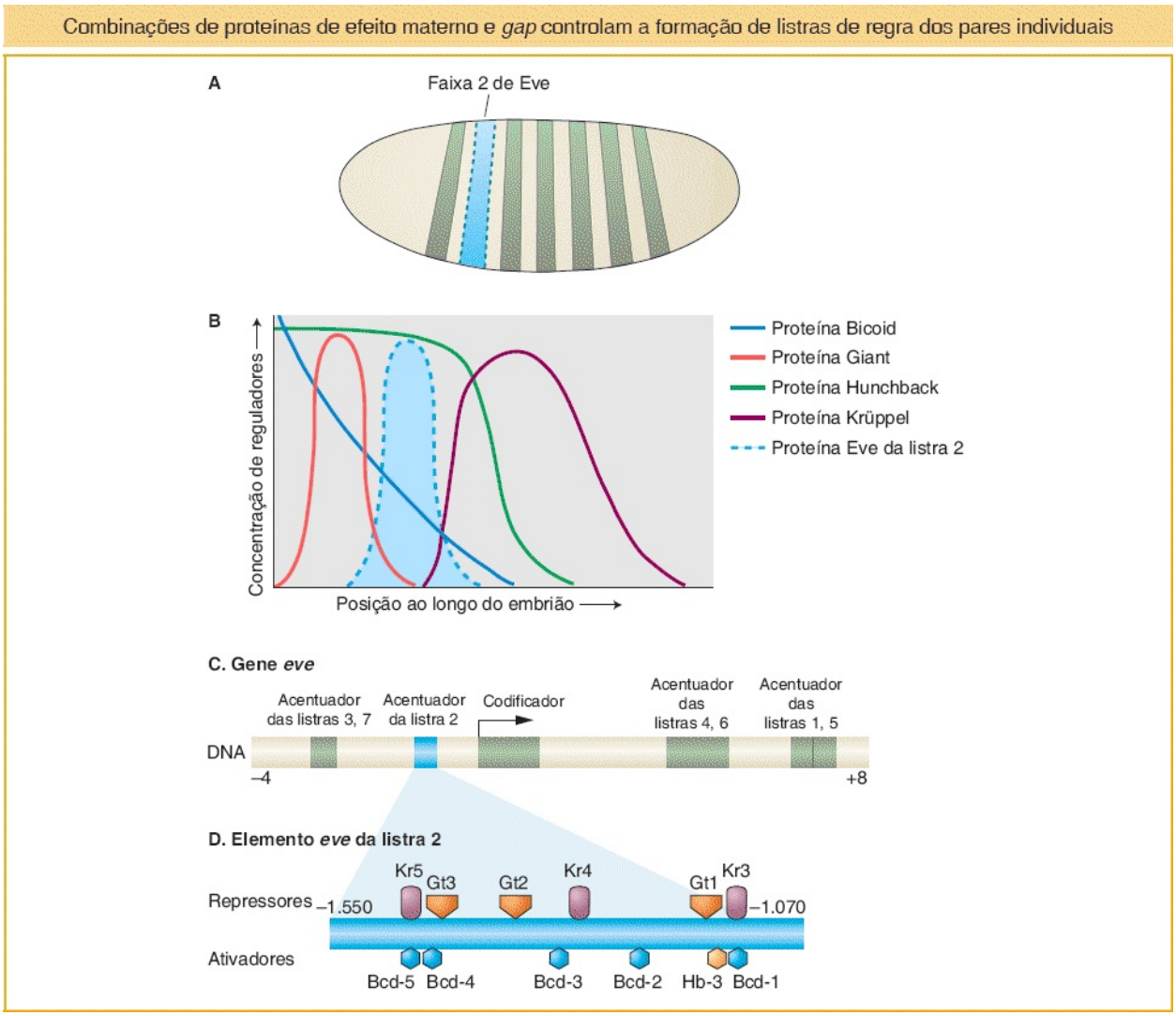


FIGURA 13.15 Regulação de uma listra de regra dos pares: controle combinatório de um elemento regulador de ação cis independente. **A.** A regulação do elemento regulador de ação cis *eve* da listra 2 controla a formação da segunda listra da expressão de *eve* no embrião inicial, apenas uma das sete listras de expressão de *eve*. **B.** A listra se forma nos domínios das proteínas Bicoid e Hunchback e na borda das proteínas gap Giant e Krüppel. Bcd e Hb são ativadores, Gt e Kr são repressores da listra. **C.** O elemento *eve* da listra 2 é apenas um de diversos elementos reguladores de ação cis do gene *eve*, cada um dos quais controlando diferentes partes da expressão de *eve*. O elemento *eve* da listra 2 engloba cerca de 1 a 1,7 kb *upstream* da unidade de transcrição de *eve*. **D.** No elemento *eve* da listra 2 existem diversos sítios de ligação para cada fator de transcrição (os repressores estão demonstrados acima do elemento e os ativadores, abaixo). A produção líquida dessa combinação de ativadores e repressores é a expressão da estreita faixa *eve*.

Especificamente, o elemento *eve* da listra 2 contém múltiplos sítios para a proteína Bicoid materna e as proteínas gap Hunchback, Giant e Krüppel (Figura 13.15 D). Análises mutacionais de diferentes combinações de sítios de ligação

revelaram que Bicoid e Hunchback ativam a expressão do elemento *eve* da faixa 2 ao longo de uma região ampla. As proteínas Giant e Krüppel são repressores que modulam os limites da listra a apenas algumas células de largura. O elemento *eve* da listra 2 atua, então, como um interruptor genético, integrando múltiplas atividades de proteínas reguladoras para produzir uma listra de três a quatro células de largura no embrião.

O padrão periódico inteiro de sete listras da expressão de *even-skipped* é a soma de diferentes conjuntos de estímulos nos elementos reguladores de ação cis em separado. Os acentuadores em relação a outras listras contêm diferentes combinações de sítios de ligação proteica.

> CONCEITO-CHAVE A regulação dos elementos reguladores de ação cis por meio de combinações de ativadores e repressores é um tema comum na regulação espacial da expressão gênica. Padrões complexos de estímulos com frequência estão integrados para produzir padrões de respostas mais simples.

Diferenciação dos segmentos | Integração dos estímulos de Hox

A atividade combinada e sequencial das proteínas de efeito materno, *gap*, regra dos pares e proteínas de polaridade do segmento estabelece o plano corporal segmentado básico do embrião e da larva. Como as diferentes identidades dos segmentos são estabelecidas pelas proteínas Hox? Esse processo apresenta dois aspectos. Primeiramente, os genes *Hox* são expressos em diferentes domínios ao longo do eixo anteroposterior. A expressão do gene *Hox* é amplamente controlada por proteínas de segmentação, especialmente proteínas gap, por meio de mecanismos que são semelhantes àqueles já descritos aqui em relação a *hunchback* e *eve* da faixa 2 (bem como a alguma regulação cruzada por parte das proteínas Hox de outros genes *Hox*). A regulação dos genes *Hox* não será considerada de modo aprofundado aqui. O segundo aspecto do controle da identidade dos segmentos por Hox é a regulação dos genes-alvo por proteínas

Hox. Examinaremos um exemplo que ilustra bem como uma característica importante do plano corporal da mosca-das-frutas é controlada por meio da integração de muitos estímulos por um único elemento regulador de ação cis.

Os membros pareados, partes da boca e antenas de *Drosophila* se desenvolvem a partir de populações inicialmente pequenas de aproximadamente 20 células em diferentes segmentos. Diferentes estruturas se desenvolvem a partir dos diferentes segmentos da cabeça e do tórax, enquanto o abdome não apresenta membros. O primeiro sinal de desenvolvimento dessas estruturas é a ativação de genes reguladores dentro de pequenos grupos de células, que são denominados *primórdios* de apêndices. A expressão do gene *Distal-less* (*Dll*) marca o início do desenvolvimento dos apêndices. Esse gene é um dos alvos-chave dos genes *Hox* e a sua função é necessária para o subsequente desenvolvimento das partes distais de cada um desses apêndices. Os pequenos grupos de células que expressam *Distal-less* surgem em diversos segmentos da cabeça e em cada um dos três segmentos torácicos, mas não no abdome (Figura 13.16 A).

Proteínas Hox reprimem a formação de apêndices no abdome

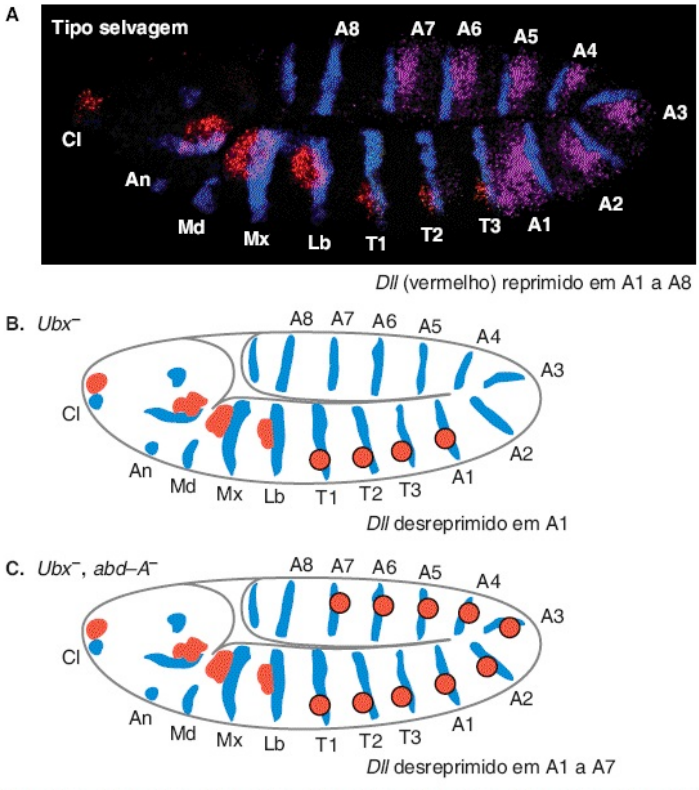


FIGURA 13.16 A ausência de membros no abdome é controlada pelos genes *Hox*. **A.** A expressão do gene *Distal-less* (*Dll*) (*vermelho*) marca a posição dos futuros apêndices, a expressão do gene *Hox Ultrabithorax* (*roxo*) marca a posição dos segmentos abdominais A1 a A7 e a expressão do gene *engrailed* (*azul*) marca a parte posterior de cada segmento. **B.** Representação esquemática de embrião *Ubx*¯ demonstrando que a expressão de *Dll* (*círculos vermelhos*) deixa de ser reprimida no segmento A1. **C.** Representação esquemática do embrião *Ubx*¯ *abd-A*¯ demonstrando que a expressão de *Dll* (*círculos vermelhos*) deixa de ser reprimida nos primeiros sete segmentos abdominais. (*A.* Fotomicrografia por Dave Kosman, Ethan Bier e Bill McGinnis; *B* e *C*. Dados de B. Gebelein, D. J. McKay e R. S. Mann, "Direct integration of Hox and Segmentation Gene Inputs During Drosophila Development", *Nature* 431, 2004, 653-659.)

Como a expressão de *Distal-less* é restrita aos segmentos mais anteriores? Por meio da repressão de sua expressão no abdome. Diversas linhas de evidências revelaram que o gene *Distal-less* é reprimido por duas proteínas Hox — as proteínas Ultrabithorax e Abdominal-A — que atuam em colaboração com duas proteínas de segmentação. Observe na Figura 13.5 que Ultrabithorax é expressa nos segmentos abdominais um a sete e Abdominal-A é expressa nos segmentos abdominais dois a sete, com sobreposição de todos os segmentos, com exceção do primeiro segmento, coberto por Ultrabithorax. Em embriões mutantes *Ultrabithorax*, a expressão de *Distal-less* se expande até o primeiro segmento abdominal (Figura 13.16 B) e em embriões duplos mutantes *Ultrabithorax/Abdominal-A*, a expressão de *Distal-less* se estende pelos primeiros sete segmentos abdominais (Figura 13.16 C), indicando que ambas as proteínas são necessárias para a repressão da expressão de *Distal-less* no abdome.

O elemento regulador de ação cis responsável pela expressão de *Distal-less* no embrião foi identificado e caracterizado em detalhes (Figura 13.17 A). Ele contém dois sítios de ligação para as proteínas Hox. Se esses dois sítios de ligação estiverem mutados de tal modo que as proteínas Hox não consigam se ligar, a expressão de *Distal-less* deixa de ser reprimida no abdome (Figura 13.17 B). Diversas proteínas adicionais colaboram com as proteínas Hox na repressão de *Distal-less*. Duas são proteínas codificadas por genes de polaridade do segmento, *Sloppy-paired* (*Slp*) e *engrailed* (*en*). As proteínas Sloppy-paired e Engrailed são expressas nas listras que marcam os compartimentos anterior e posterior de cada segmento, respectivamente. Cada proteína também se liga ao

elemento regulador de ação cis de *Distal-less*. Quando o sítio de ligação Sloppy-paired está mutado no elemento regulador de ação cis, a expressão do gene repórter deixa de ser reprimida nos compartimentos anteriores dos segmentos abdominais (Figura 13.17 C). Quando o sítio de ligação de Engrailed está mutado, a expressão do repórter deixa de ser reprimida nos compartimentos posteriores de cada segmento abdominal (Figura 13.17 D). E quando os sítios de ligação em relação a ambas as proteínas estão mutados, a expressão do gene repórter deixa de ser reprimida em ambos os compartimentos de cada segmento abdominal, assim como quando os sítios de ligação Hox estão mutados (Figura 13.17 C). Duas outras proteínas, denominadas Extradenticle e Homothorax, que são amplamente expressas em cada segmento, também se ligam ao elemento regulador de ação cis de *Distal-less* e são necessárias para a repressão da transcrição no abdome (Figura 13.17 F).

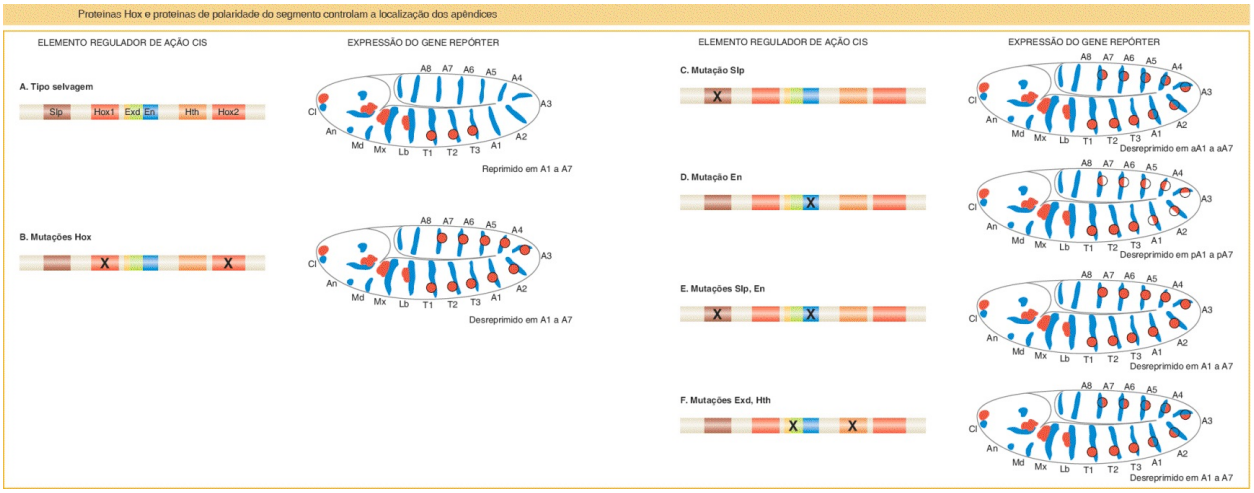


FIGURA 13.17 Integração dos estímulos das proteínas Hox e de segmentação por meio de um elemento regulador de ação cis. **A.** Um elemento regulador de ação cis do gene *Dll* regula a repressão da expressão de *Dll* no abdome por meio de um conjunto de fatores de transcrição (*esquerda*). A expressão de *Dll* (*vermelho*) se estende até o tórax, mas não para dentro do abdome em um embrião do tipo selvagem (*direita*). **B** a **F.** Mutações nos respectivos sítios de ligação demonstram desrepressão da expressão de *Dll* em diversos padrões no abdome. Os sítios de ligação são: Slp, Sloppy-paired; Hox1 e Hox2, Ultrabithorax e Abdominal-A; Exd, Extradenticle; En, Engrailed; Hth, Homothorax. (Dados de B. Gebelein, D. J. McKay e R. S. Mann, "Direct Integration of Hox and Segmentation Gene Inputs During Drosophila Development", *Nature* 431, 2004, 653-659.)

Portanto, em conjunto, duas proteínas Hox e quatro outros fatores de transcrição se ligam a um trecho de 57 pares de bases e atuam em conjunto para reprimir a expressão *Distal-less* e, portanto, a formação de apêndices no abdome. A repressão da expressão *Distal-less* é uma clara demonstração de como as proteínas Hox regulam a identidade dos segmentos e o número de estruturas corporais reiteradas. Também é uma boa ilustração de como diversos estímulos reguladores atuam de modo combinatório em elementos reguladores de ação cis. Nesse caso, a presença de sítios de ligação Hox não é suficiente para a repressão da transcrição: são necessárias interações colaborativas e cooperativas entre diversas proteínas para reprimir totalmente a expressão gênica no abdome.

> CONCEITO-CHAVE A regulação combinatória e cooperativa da transcrição gênica impõe maior especificidade sobre os padrões espaciais de expressão gênica e possibilita a sua maior diversidade.

Embora a diversidade evolutiva não tenha sido explicitamente abordada neste capítulo, a presença de múltiplos elementos reguladores de ação cis independentes para cada gene *toolkit* apresenta profundas implicações para a evolução da forma. Especificamente, a modularidade desses elementos possibilita alterações em um aspecto da expressão gênica independente de outras funções dos genes. A evolução da regulação gênica desempenha um papel importante na evolução do desenvolvimento e da morfologia. Retornaremos a esse tópico no Capítulo 20.

13.5 Regulação pós-transcricional da expressão gênica no desenvolvimento

Embora a regulação da transcrição seja um meio importante para restringir a expressão dos produtos gênicos a áreas definidas durante o desenvolvimento, ela absolutamente não é o meio exclusivo para tanto. A recomposição alternativa do

RNA também contribui para a regulação gênica, assim como a regulação da tradução do mRNA por proteínas e microRNA (miRNA). Em cada caso, sequências reguladoras no RNA são reconhecidas — por meio de fatores de recomposição, proteínas de ligação ao mRNA ou miRNA — e regulam a estrutura do produto proteico, sua quantidade ou a localização na qual a proteína é produzida. Veremos um exemplo de cada tipo de interação reguladora no nível do RNA.

Recomposição do RNA e determinação do sexo em Drosophila

Uma decisão fundamental no desenvolvimento de organismos com reprodução sexuada é a especificação do sexo. Em animais, o desenvolvimento de muitos tecidos segue vias diferentes, dependendo do sexo do animal em questão. Em *Drosophila*, foram identificados muitos genes que regulam a *determinação do sexo* por meio da análise de fenótipos mutantes nos quais a identidade sexual está alterada ou é ambígua.

O gene *doublesex* (*dsx*) desempenha um papel central na regulação da identidade sexual do tecido somático (linhagem não germinativa). Mutações nulas em *dsx* causam o desenvolvimento de machos e fêmeas como *intersexos* intermediários, que perderam as diferenças distintas entre os tecidos masculinos e femininos. Embora a função de *dsx* seja necessária em ambos os sexos, diferentes produtos gênicos são gerados pelo *locus* em diferentes sexos. Nos machos, o produto é uma isoforma específica e mais longa, Dsx^{M}, que contém uma região C-terminal única de 150 aminoácidos não observada na isoforma específica feminina Dsx^{F}, que, em vez disso, contém uma sequência única de 30 aminoácidos na carboxi-terminal. Cada forma da proteína Dsx é um fator de transcrição de ligação ao DNA que aparentemente liga-se às mesmas sequências de DNA. Entretanto, as atividades das duas isoformas diferem: Dsx^{F} ativa determinados genes-alvo nas fêmeas que Dsx^{M} reprime nos machos.

As formas alternativas da proteína Dsx são geradas por meio da recomposição alternativa do transcrito de RNA primário de *dsx*. Portanto, nesse caso, a escolha dos sítios de recomposição deve ser regulada para produzir mRNA finais que codificam proteínas diferentes. Os diversos fatores genéticos que influenciam a

expressão de Dsx e a determinação do sexo foram identificados por meio de mutações que afetam o fenótipo sexual.

Um regulador-chave é o produto do gene *transformer* (*tra*). Enquanto mutações nulas em *tra* não apresentam efeitos nos machos, moscas-fêmeas XX que apresentam mutações em *tra* são transformadas no fenótipo masculino. A proteína Tra é um fator de recomposição alternativa que afeta as opções de recomposição no transcrito de RNA de *dsx*. Na presença de Tra (e uma proteína correlata Tra2), ocorre uma recomposição que incorpora o éxon 4 do gene *dsx* ao transcrito dsx^{F} maduro (Figura 13.18), mas não os éxons 5 e 6. Os machos apresentam ausência da proteína Tra; assim, essa recomposição não ocorre e os éxons 5 e 6 são incorporados ao transcrito dsx^{M}, mas não o éxon 4.

A proteína Tra explica como formas alternativas de Dsx são expressas, mas como a expressão da própria Tra é regulada para diferir em fêmeas e machos? O próprio RNA de *tra* é recomposto alternativamente. Nas fêmeas, está presente um fator de recomposição codificado pelo gene *Sex-lethal* (*Sxl*). Esse fator de recomposição se liga ao RNA de *tra* e evita o evento de recomposição que, de outro modo, incorporaria um éxon que contém um códon de fim. Nos machos, nenhuma proteína Tra é produzida, tendo em vista que esse códon de fim está presente.

A produção da proteína Sex-lethal, por sua vez, é regulada pela recomposição do RNA e por fatores que alteram o nível de transcrição. O nível de transcrição de *Sxl* é regulado por ativadores no cromossomo X e repressores nos autossomos. Nas fêmeas, a ativação de *Sxl* prevalece e é produzida a proteína Sxl, a qual regula a recomposição do RNA de *tra* e realiza um *feedback* para regular a recomposição do próprio RNA de *Sxl*. Nas fêmeas, um códon de fim é recomposto, de modo que a produção da proteína Sxl possa continuar. Entretanto, nos machos, nos quais não há proteína Sxl, o códon de fim ainda está presente no transcrito de RNA de *Sxl* não recomposto e nenhuma proteína Sxl pode ser produzida.

Essa cascata de recomposição de RNA sexo-específico em *D. melanogaster* ilustra um modo pelo qual o genótipo do cromossomo sexual leva à expressão de diferentes formas de proteínas reguladoras em um sexo e não no outro.

Curiosamente, a regulação genética da determinação do sexo difere muito entre as espécies de animais, pois o genótipo sexual pode levar à expressão diferencial de genes reguladores por meio de vias claramente diferentes. Entretanto, as proteínas relacionadas com a Dsx desempenham papéis na diferenciação sexual em uma ampla variedade de animais, incluindo seres humanos. Portanto, embora existam muitos modos para gerar a expressão diferencial de fatores de transcrição, uma família de proteínas semelhantes aparenta ser a base de uma grande parte da diferenciação sexual.

Regulação da tradução do mRNA e linhagem celular em C. elegans

Em muitas espécies de animais, o desenvolvimento inicial do embrião envolve a repartição de células ou grupos de células em linhagens discretas que darão origem a tecidos distintos no adulto. Esse processo é mais bem-compreendido no verme nematódeo *C. elegans*, no qual o animal adulto é composto por apenas aproximadamente 1.000 células somáticas (um terço das quais é composto de células nervosas) e um número semelhante de células germinativas na gônada. A estrutura simples, o ciclo de vida rápido e a transparência do *C. elegans* o tornaram um modelo poderoso para a análise do desenvolvimento (ver Organismo-modelo, *C. elegans*, adiante). Todas as linhagens celulares desse animal foram mapeadas em uma série de estudos liderados por John Sulston no laboratório Medical Research Council (MRC), em Cambridge, na Inglaterra. Triagens genéticas sistemáticas em relação a mutações que modificam ou estendem linhagens celulares proporcionaram uma recompensa de informações a respeito do controle genético de decisões das linhagens. A genética de *C. elegans* tem sido especialmente importante para a compreensão do papel da regulação pós-transcricional no nível do RNA e aqui examinaremos dois mecanismos: (1) o controle da tradução por meio de proteínas de ligação ao mRNA e (2) o controle da expressão gênica por miRNA.

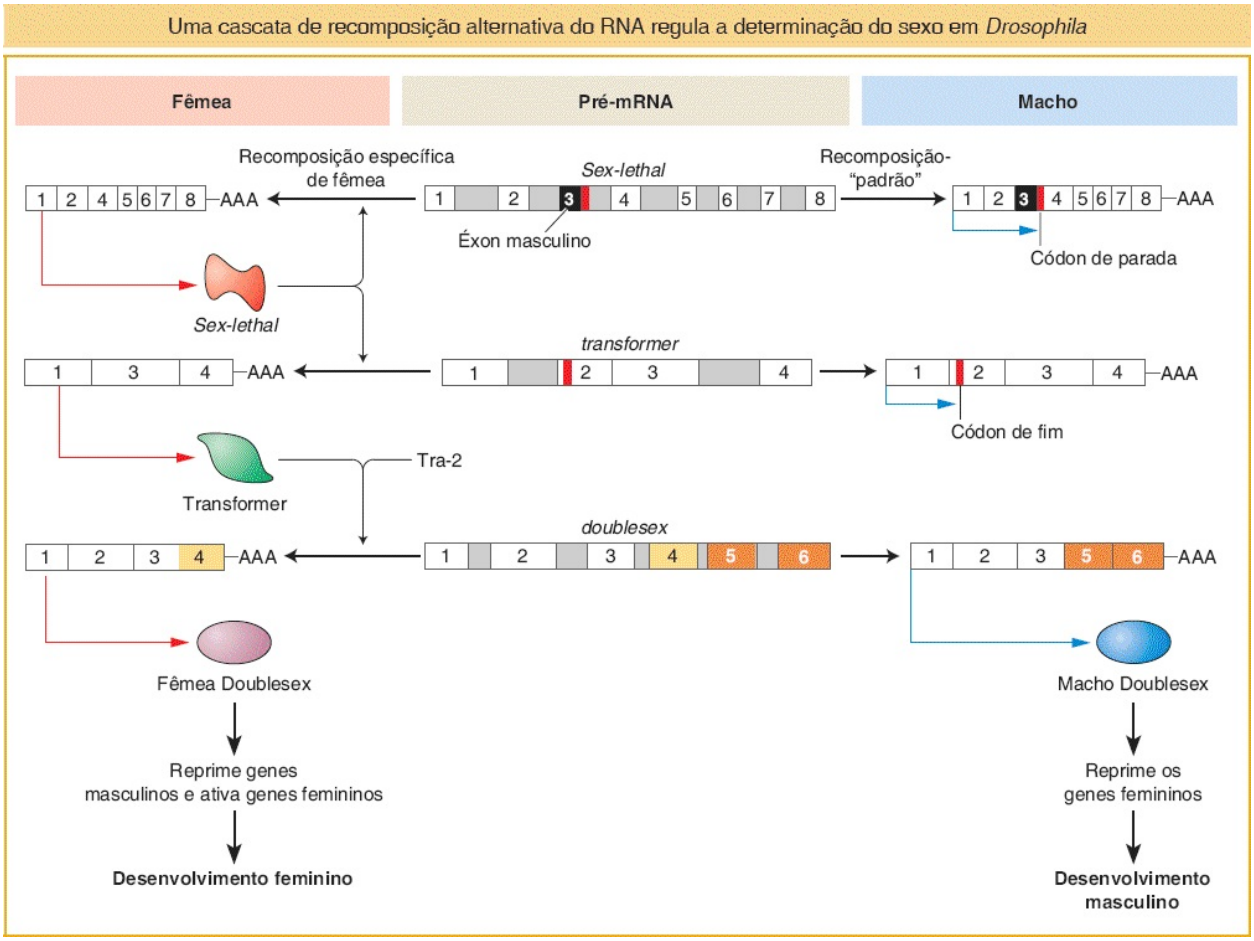


FIGURA 13.18 Três pré-mRNA de importantes genes de determinação do sexo em *Drosophila* são recompostos alternativamente. A via específica das fêmeas está demonstrada à esquerda e a via específica dos machos está demonstrada à direita. Os pré-mRNA são idênticos em ambos os sexos e estão demonstrados na parte intermediária. No macho, existem códons de fim nos mRNA *Sex-lethal* e *transformer*, que encerram a tradução. Essas sequências são removidas por recomposição para produzir proteínas funcionais na fêmea. As proteínas Transformer e Tra-2 em seguida recompõem o pré-mRNA de *doublesex* da fêmea para produzir a isoforma específica da fêmea da proteína Dsx, que difere da isoforma específica do macho pela recomposição alternativa de diversos éxons.

Controle da tradução no embrião inicial

Primeiramente veremos como uma linhagem celular tem início. Após duas divisões celulares, o embrião de *C. elegans* contém quatro células, denominadas blastômeros. Cada célula dará início a uma linhagem distinta e os descendentes das linhagens separadas apresentarão destinos diferentes. Já nesse estágio, são observadas diferenças nas proteínas presentes nos quatro blastômeros. Pelo que

já aprendemos, muitas dessas proteínas são proteínas toolkit que determinam quais genes serão expressos nas células descendentes. Entretanto, o que é surpreendente é que os mRNA que codificam algumas proteínas toolkit do verme estejam presentes em *todas* as células do embrião inicial. Entretanto, em uma célula específica, apenas alguns desses mRNA serão traduzidos em proteínas. Portanto, no embrião de *C. elegans*, a regulação pós-transcricional é crítica para a adequada especificação dos destinos celulares iniciais. Durante a primeira divisão celular, a polaridade dentro do zigoto leva à repartição de moléculas reguladoras para células embrionárias específicas. Por exemplo, o gene *glp-1* codifica uma proteína de receptor transmembrana (relacionada com o receptor Notch de moscas e outros animais). Embora o mRNA de *glp-1* esteja presente em todas as células no estágio de quatro células, a proteína GLP-1 é traduzida apenas nas duas células anteriores Aba e ABp (Figura 13.19 A). Essa expressão localizada de GLP-1 é crítica para o estabelecimento de destinos distintos. As mutações que eliminam a função de *glp-1* no estágio de quatro células alteram os destinos dos descendentes de ABp e ABa.

ORGANISMO-MODELO *Caenorhabditis elegans*

O nematódeo Caenorhabditis elegans como um modelo para decisões do destino de linhagens celulares

Nos últimos 20 anos, estudos do verme nematódeo *Caenorhabditis elegans* (ver Diagrama 1) avançaram muito a nossa compreensão sobre o controle genético das decisões de linhagens celulares. A transparência e a estrutura simples desse animal levaram Sydney Brenner a avançar a sua utilização como um organismo-modelo. O verme adulto contém aproximadamente 1.000 células somáticas e pesquisadores, liderados por John Sulston, mapearam cuidadosamente a série inteira de decisões de células somáticas que produzem o animal adulto.

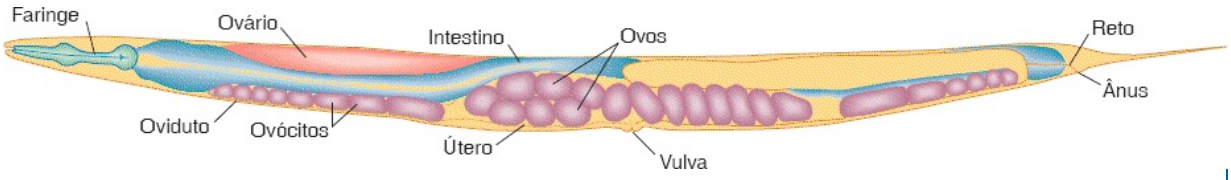


Diagrama 1 *Caenorhabditis elegans* hermafrodita adulto, demonstrando diversos órgãos.

Algumas das decisões de linhagens, tais como a formação da vulva, têm sido modelos-chave das assim denominadas *interações indutivas* no desenvolvimento, nas quais a sinalização entre as células induz as alterações no destino celular e a formação dos órgãos (ver Diagrama 2). Triagens genéticas exaustivas identificaram muitos componentes que participam na sinalização e na transdução de sinal na formação da vulva.

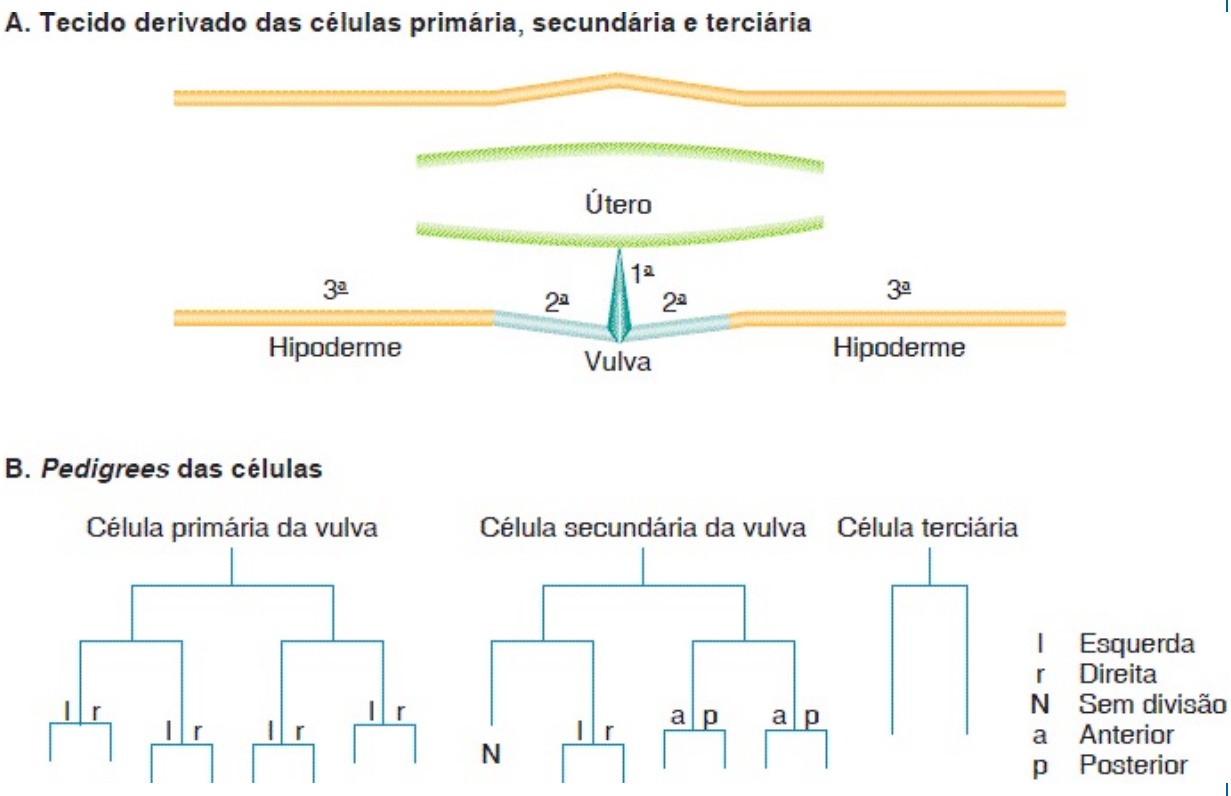


Diagrama 2 Produção das linhagens celulares da vulva. A. As partes da anatomia vulvar que estão ocupadas pelas assim denominadas células primárias (1ª), secundárias (2ª) e terciárias (3ª). B. As linhagens ou os *pedigrees* das células primárias, secundárias e terciárias são distinguidas por

meio de seus padrões de divisão celular.

Em relação a algumas das divisões celulares embrionárias e larvais, em particular aquelas que contribuirão para o sistema nervoso do verme, uma célula genitora dá origem a uma progênie de duas células, uma das quais em seguida é submetida à morte celular programada. A análise de mutantes nos quais a morte celular programada é aberrante, liderada por Robert Horvitz, revelou muitos componentes das vias da morte celular programada comuns à maior parte dos animais. Sydney Brenner, John Sulston e Robert Horvitz compartilharam o Prêmio Nobel de Fisiologia ou Medicina de 2002 por seu trabalho pioneiro com base em *C. elegans*.

A GLP-1 está localizada nas células anteriores em virtude da repressão de sua tradução nas células posteriores. A repressão da tradução de GLP-1 necessita de sequências na 3′ UTR do mRNA de *glp-1* — especificamente, uma região de 61 nucleotídios denominada região de controle espacial (SCR). A importância da SCR foi demonstrada por meio da ligação do mRNA transcrito de genes repórteres a diferentes variantes de SCR. A deleção dessa região ou a mutação de sítios-chave dentro dela leva à expressão do gene repórter em todos os quatro blastômeros do embrião inicial (Figura 13.19 B).

Com base no modo como vimos o controle da transcrição, podemos imaginar que uma ou mais proteínas se ligam à SCR para reprimir a tradução do mRNA de *glp-1*. Para identificar essas proteínas repressoras, pesquisadores isolaram proteínas que se ligam à SCR. Uma proteína, GLD-1, liga-se especificamente a uma região da SCR. Além disso, a proteína GLD-1 está enriquecida em blastômeros posteriores, justamente onde a expressão de *glp-1* é reprimida. Finalmente, quando a expressão de GLD-1 é inibida pelo uso de RNA de interferência, a proteína GLP-1 é expressa nos blastômeros posteriores (Figura 13.19 C). Essa evidência sugere que a GLD-1 é uma proteína repressora da tradução que controla a expressão de *glp-1*.

Proteínas de ligação ao mRNA reprimem a tradução do mRNA para determinar as linhagens celulares

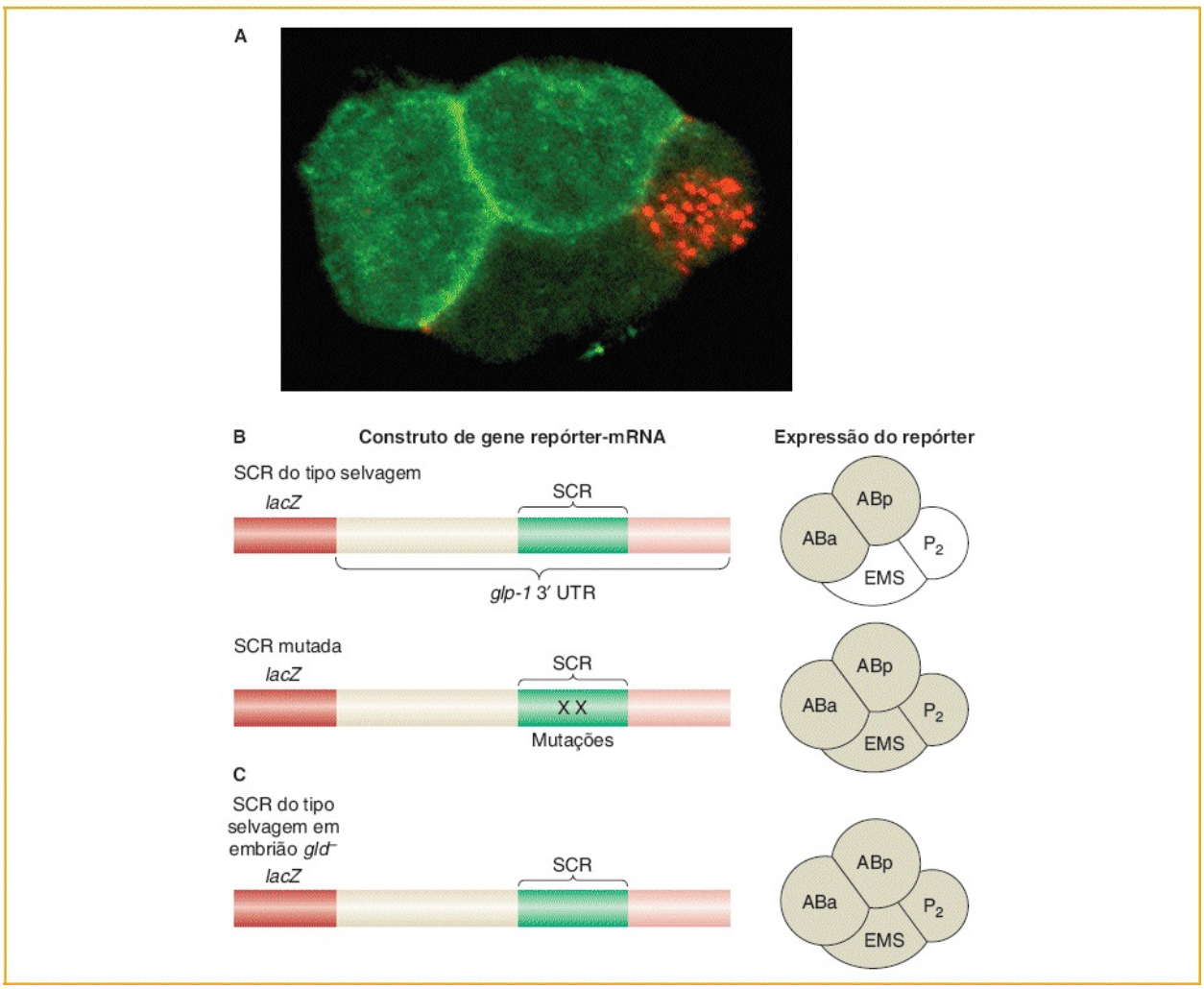


FIGURA 13.19 Regulação da tradução e decisões de linhagem celular no embrião inicial de *C. elegans*. **A.** No estágio de quatro células do embrião de *C. elegans*, a proteína GLP-1 é expressa em duas células anteriores (*verde brilhante*), mas não em outras células. A tradução do mRNA de *glp-1* é regulada pela proteína GLD-1 nas células posteriores. **B.** A fusão de 3′ UTR *glp-1* ao gene repórter *lacZ* leva à expressão do repórter nas células ABa e ABp do estágio de quatro células do embrião de *C. elegans* (*sombreado, direita*). Mutações em sítios de ligação de GLD-1 na região de controle espacial (SCR) causam desrepressão da tradução nas linhagens EMS e P_2, assim como a perda da função de *gld* (**C**). (*A*. Cortesia de Thomas C. Evans, University of Colorado Anschutz Medical Campus.)

A regulação espacial da tradução de GLP-1 nada mais é do que um exemplo do controle da tradução no desenvolvimento ou pela GLD-1. Muitos outros mRNA são regulados em relação à tradução e a GLD-1 se liga a outros mRNA-alvo nas células embrionárias e da linhagem germinativa.

> CONCEITO-CHAVE As proteínas de ligação ao RNA sequência-específicas atuam por meio de sequências de RNA de ação cis para regular o padrão espacial da tradução da proteína.

Controle da cronologia do desenvolvimento por miRNA em C. elegans e outras espécies

O desenvolvimento é um processo ordenado tanto temporal quanto espacialmente. Quando os eventos ocorrem é tão importante quanto onde. As mutações nos genes *heterocrônicos* de *C. elegans* têm sido fontes de percepção sobre o controle da cronologia do desenvolvimento. Mutações nesses genes alteram a ocasião dos eventos na especificação do destino celular, fazendo com que os referidos eventos sejam reiterados ou omitidos. A investigação detalhada dos produtos de genes heterocrônicos levou à descoberta de um mecanismo totalmente inesperado em relação à regulação da expressão gênica, por meio de microRNA.

Entre os primeiros membros dessa classe de moléculas reguladoras descobertas em *C. elegans* está o RNA produzido pelo gene *let-7*. Tal gene regula a transição dos destinos celulares dos últimos estágios larvais para o adulto. Em mutantes *let-7*, por exemplo, os destinos das células larvais são reiterados no estágio adulto (Figura 13.20 A). Contrariamente, o aumento de dosagem do gene *let-7* causa especificação precoce dos destinos adultos nos estágios larvais.

O gene *let-7* não codifica uma proteína. Em vez disso, ele codifica um RNA final, de 22 nucleotídios, regulado temporalmente, o qual é processado a partir de um precursor de aproximadamente 70 nucleotídios. O RNA final é complementar às sequências nas regiões 3′ não traduzidas de uma diversidade de genes regulados durante o desenvolvimento e a ligação de miRNA a essas sequências impede a tradução desses transcritos gênicos. Um desses genes-alvo, *lin-41*, também afeta a transição de larva para adulto. Os mutantes de *lin-41* causam a especificação precoce dos destinos celulares adultos, sugerindo que o efeito da hiperexpressão de *let-7* ocorre, pelo menos em parte, em virtude de um efeito na expressão de *lin-41*. O mRNA de *let-7* se liga ao RNA de *lin-41 in vitro* em

diversos sítios complementares imperfeitos (Figura 13.20 B).

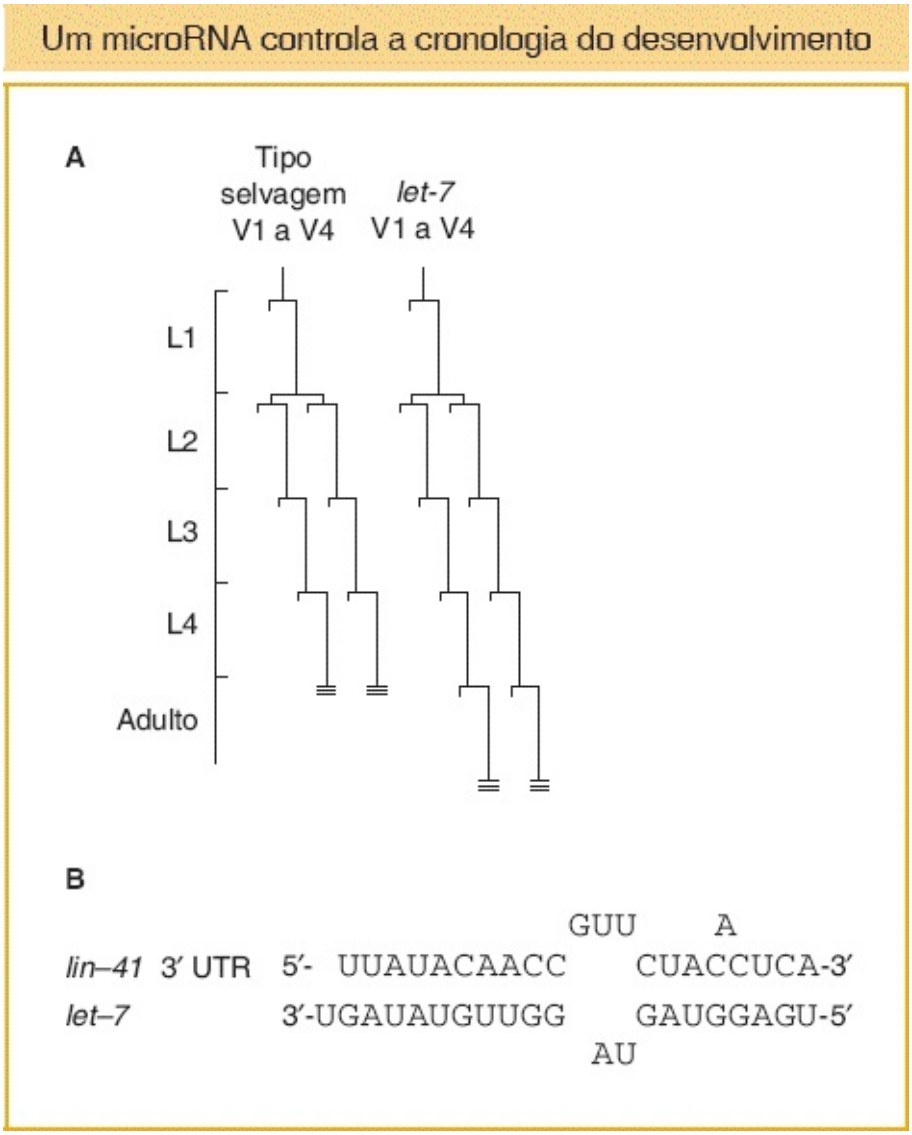


FIGURA 13.20 Normalmente, *C. elegans* se desenvolve em um adulto após quatro estágios larvais e as linhagens celulares hipodérmicas concluem o seu desenvolvimento em L4 (linhagens eclodidas nas extremidades das linhagens V1 a V4). **A.** Em mutantes *let-7*, a transição do estágio larvar L4 para adulto é adiada e as linhagens celulares de células hipodérmicas laterais (V) são reiteradas. **B.** *let-7* codifica um miRNA que é complementar a sequências em 3′ UTR do mRNA de *lin-41*.

O papel dos miRNA no desenvolvimento de *C. elegans* se estende muito além de *let-7*. Foram identificadas diversas centenas de miRNA e foi demonstrado que muitos genes-alvo são regulados pelo miRNA. Além disso, a descoberta dessa classe de RNA regulador provocou a busca pelos referidos genes em outros

genomas e, em geral, centenas de genes de miRNA candidatos foram detectados em genomas animais, incluindo aquele de seres humanos.

De modo razoavelmente surpreendente, o gene de miRNA *let-7* é amplamente conservado e encontrado nos genomas de *Drosophila*, ascídias, moluscos, anelídeos e vertebrados (incluindo seres humanos). O gene *lin-41* também é conservado e evidências sugerem que a interação reguladora de *let-7* e *lin-41* também controla a cronologia dos eventos no desenvolvimento de outras espécies.

As descobertas da regulação por miRNA dos genes de desenvolvimento e do escopo do repertório de miRNA são razoavelmente recentes. Geneticistas e outros biólogos estão consideravelmente entusiasmados a respeito dos papéis dessa classe de moléculas reguladoras no desenvolvimento e na fisiologia, levando a uma área muito vigorosa e acelerada de novas pesquisas.

13.6 De moscas a dedos, penas e placas do assoalho | Os muitos papéis dos genes toolkit

Vimos que as proteínas toolkit e os RNA reguladores apresentam múltiplos papéis no desenvolvimento. Por exemplo, relembre que, na mosca, a proteína Ultrabithorax reprime a formação dos membros no abdome e promove o desenvolvimento das asas traseiras no tórax. De modo semelhante, a Sloppy-paired e a Engrailed participam na geração da organização segmentar básica do embrião e colaboram com as proteínas Hox para suprimir a formação dos membros. Esses papéis são apenas alguns dos muitos papéis desempenhados por esses genes *toolkit* em todo o período do desenvolvimento da mosca. A maior parte dos genes *toolkit* atua em mais de uma ocasião e em mais de um local e a maior parte pode influenciar a formação ou a padronização de muitas estruturas distintas que são formadas em diferentes partes do corpo da larva ou do adulto. Aqueles que regulam a expressão gênica podem regular diretamente dezenas, até centenas, de genes diferentes. A função de uma proteína toolkit (ou RNA) individual quase sempre é dependente do contexto, motivo pelo qual a analogia

com uma caixa de ferramentas talvez seja bastante adequada. Assim como a caixa de ferramentas de um carpinteiro, um conjunto comum de ferramentas pode ser utilizado para produzir muitas estruturas.

Para ilustrar esse princípio mais claramente, observaremos o papel de uma proteína toolki*t* no desenvolvimento de muitas características de vertebrados, incluindo características presentes em seres humanos. Essa proteína toolkit é homóloga nos vertebrados àquela do gene *hedgehog* de *Drosophila*. O gene *hedgehog* foi identificado pela primeira vez por Nüsslein-Volhard e Wieschaus como um gene de polaridade de segmento. Ele foi caracterizado como codificador de uma proteína de sinalização secretada por células em *Drosophila*.

Na medida em que cresciam as evidências de que os genes *toolkit* são comuns a diferentes filos de animais, a descoberta e a caracterização de genes *toolkit* da mosca, tais como o *hedgehog*, tornaram-se um trampolim comum para a caracterização de genes em outros grupos, particularmente em vertebrados. A clonagem de genes homólogos com base na similaridade de sequências (ver Capítulo 14) foi uma via rápida para a identificação dos genes *toolkit* nos vertebrados. A aplicação dessa estratégia para o gene *hedgehog* ilustra o poder e as recompensas da utilização da homologia para descobrir genes importantes. Diversos homólogos distintos de *hedgehog* foram isolados de peixe-zebra, camundongos, galinhas e seres humanos. No espírito extravagante da nomenclatura do gene de *Drosophila*, os três homólogos de vertebrados foram denominados *Sonic hedgehog* (em referência ao personagem de *videogame*), *Indian hedgehog* e *Desert hedgehog*.

Um dos primeiros meios para caracterizar os possíveis papéis desses genes no desenvolvimento foi examinar onde eles são expressos. Observou-se que o *Sonic hedgehog* (*Shh*) é expresso em diversas partes do desenvolvimento de galinhas e outros vertebrados. Mais intrigante foi a sua expressão na parte posterior dos brotos dos membros em desenvolvimento (Figura 13.21 A). Sabe-se há décadas que essa parte dos brotos dos membros é a *zona de polarização de atividade* (ZPA), tendo em vista que é um organizador responsável pelo estabelecimento da polaridade anteroposterior do membro e de seus dedos (Figura 13.21 B). Para testar se o *Shh* pode desempenhar um papel na função da ZPA, Cliff Tabin e seus

colegas na Harvard Medical School fizeram com que a proteína Shh fosse expressa na região *anterior* dos brotos dos membros de galinhas em desenvolvimento. Eles observaram o mesmo efeito do transplante de ZPA — a indução de dedos extras com polaridade reversa. Seus resultados foram uma evidência chocante de que Shh era o morfógeno produzido pela ZPA que há muito tempo se procurava.

O gene *toolkit Sonic hedgehog* apresenta múltiplos papéis

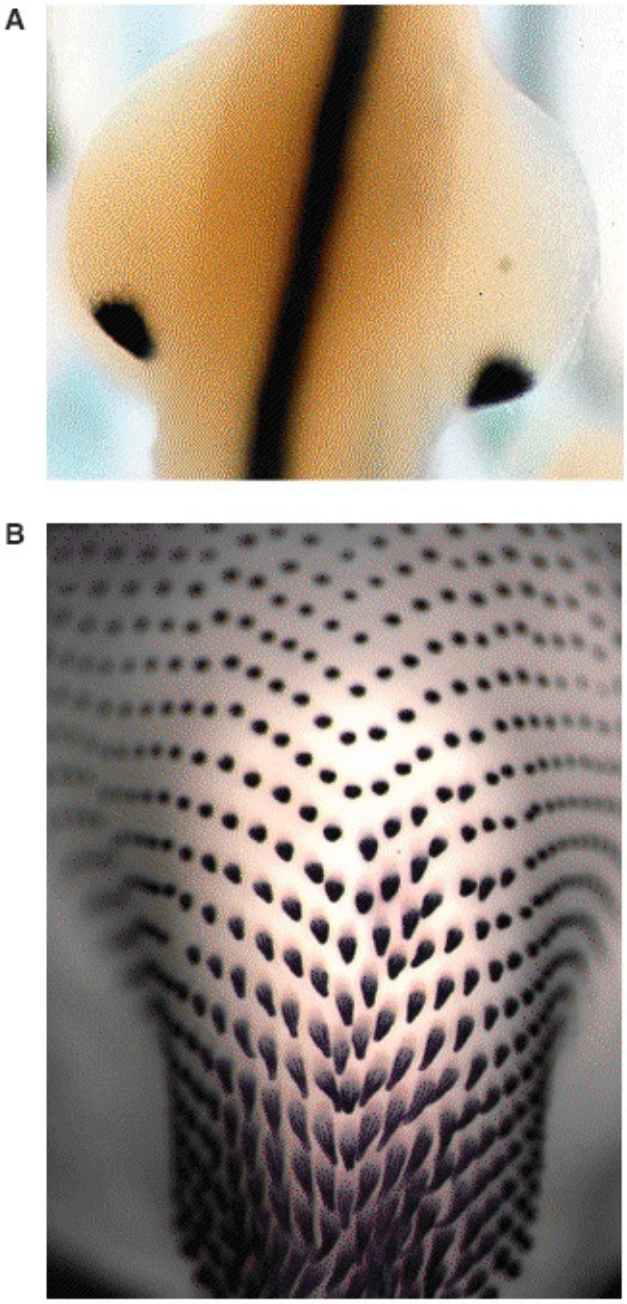


FIGURA 13.21 O gene *Shh* é expresso em muitas partes diferentes do embrião de galinha em desenvolvimento (*indicado pela coloração escura*), incluindo (**A**) a zona de atividade de polarização em cada um dos dois brotos dos membros em desenvolvimento e no longo tubo neural e (**B**) os brotos das penas em desenvolvimento. O mRNA de *Shh* é visualizado por meio de hibridização *in situ*. (As fotomicrografias são cortesia de: *A.* Cliff Tabin; *B.* Dr. John Fallon, University of Wisconsin, e Matthew Harris, Harvard Medical

School, Department of Genetics.)

A Shh também é expressa em outros padrões intrigantes em galinhas e outros vertebrados. Por exemplo, Shh é expressa em brotos de penas em desenvolvimento, onde desempenha um papel no desenvolvimento do padrão e da polaridade da formação das penas (ver Figura 13.21 B). A Shh também é expressa no tubo neural em desenvolvimento de embriões de vertebrados, em uma região denominada *placa do assoalho* (ver Figura 13.21 A). Experimentos subsequentes demonstraram que a sinalização de Shh a partir dessas células da placa do assoalho é crítica para a subdivisão dos hemisférios cerebrais e para a subdivisão do olho em desenvolvimento nos lados esquerdo e direito. Quando a função do gene *Shh* é eliminada por uma mutação no camundongo, esses hemisférios e essas regiões do olho não se separam e o embrião resultante é *ciclópico*, com um olho central e um único prosencéfalo (ele também não tem membros).

Os papéis dramáticos e diversos de *Shh* são um exemplo surpreendente dos diferentes papéis desempenhados por genes *toolkit* em diferentes locais e ocasiões no desenvolvimento. Os resultados da sinalização de Shh são diferentes em cada caso: a via de sinalização de Shh induzirá a expressão de um conjunto de genes no membro em desenvolvimento, um conjunto diferente no broto das penas e ainda outro conjunto na placa do assoalho. Como diferentes tipos celulares e tecidos são capazes de responder de modo diferente à mesma molécula sinalizadora? O desfecho da sinalização de Shh depende do contexto proporcionado pelos outros genes *toolkit* que estão atuando ao mesmo tempo.

> CONCEITO-CHAVE A maior parte dos genes *toolkit* apresenta múltiplos papéis em diferentes tecidos e tipos celulares. A especificidade de sua ação é determinada pelo contexto proporcionado por outros genes *toolkit* que atuam em uma combinação com eles.

13.7 Desenvolvimento e doença

A descoberta da *toolkit* para o desenvolvimento de moscas, vertebrados e seres humanos também apresentou um profundo efeito sobre o estudo da base genética das doenças humanas, em particular defeitos congênitos e câncer. Foi identificado um grande número de mutações nos genes *toolkit*, que afetam o desenvolvimento e a saúde humana. Aqui, enfocaremos apenas alguns exemplos que ilustram como a compreensão sobre a função e a regulação dos genes em modelos animais foi traduzida em melhor compreensão da biologia humana.

Polidactilia

Uma síndrome relativamente comum em seres humanos é o desenvolvimento de dedos extras, parciais ou completos, nas mãos e nos pés. Essa condição, denominada *polidactilia*, surge em aproximadamente 5 a 17 de cada 10.000 nascimentos vivos. Nos casos mais dramáticos, a condição está presente nas mãos e nos pés (Figura 13.22). A polidactilia ocorre amplamente em todos os vertebrados — em gatos, galinhas, camundongos e outras espécies.

A descoberta do papel da Shh na padronização dos dedos levou os geneticistas a investigarem se o gene *Shh* estava alterado em seres humanos com polidactilia e outras espécies. De fato, determinadas mutações de polidactilia são mutações no gene *Shh*. As mutações não estão na região codificadora do gene *Shh*; em vez disso, se encontram em um elemento regulador de ação cis, longe da região codificadora, que controla a expressão de *Shh* no broto do membro em desenvolvimento. Os dedos extras são induzidos pela expressão de *Shh* em uma parte do membro na qual o gene normalmente não é expresso. Mutações em elementos reguladores de ação cis apresentam duas propriedades importantes que são distintas das mutações nas regiões codificadoras. Primeiramente, tendo em vista que elas afetam a regulação em cis, os fenótipos com frequência são dominantes. Em segundo lugar, tendo em vista que apenas um dos diversos elementos reguladores de ação cis pode estar afetado, as outras funções do gene podem ser completamente normais. A polidactilia pode ocorrer sem quaisquer problemas colaterais de desenvolvimento. Entretanto, as mutações codificadoras em *Shh* contam uma história diferente, conforme veremos na próxima seção.

FIGURA 13.22 Esta pessoa apresenta seis dedos em cada mão e sete dedos em cada pé em virtude de uma mutação reguladora no gene *Sonic hedgehog*. (Cortesia de Dr. Robert Hill, MRC Human Genetics Unit, Edinburgo, Escócia; de L. A. Lettice *et al.*, "Disruption of a Long-Range Cis-Acting Regulator for Shh Causes Preaxial Polydactyly", *Proc. Natl. Acad. Sci. USA* 99, 2002, 7548 © 2002 National Academy of Sciences, USA.)

Holoprosencefalia

Também foram identificadas mutações na região codificadora do *Shh* humano. As consequentes alterações na proteína Shh estão associadas a uma síndrome denominada *holoprosencefalia*, na qual ocorrem anormalidades no tamanho do cérebro, na formação do nariz e em outras estruturas da linha média. Essas anormalidades aparentam ser correspondentes menos graves dos defeitos do desenvolvimento observados em camundongos mutantes *Shh* homozigotos. De fato, as crianças afetadas consultadas nas clínicas são heterozigotas. Uma cópia de um gene *Shh* normal parece ser insuficiente para o desenvolvimento da linha média normal (o gene é *haploinsuficiente*). Fetos humanos homozigotos para mutações de perda da função em *Shh* muito provavelmente morrem na gestação com defeitos mais graves.

A holoprosencefalia não é causada exclusivamente por mutações em *Shh*. A Shh é um ligante em uma via de transdução de sinal. Conforme se pode esperar, mutações nos genes que codificam outros componentes da via afetam a eficiência da sinalização de Shh e também estão associadas à holoprosencefalia. Diversos componentes da via Shh humana foram identificados pela primeira vez como

homólogos de membros da via na mosca, demonstrando, novamente, a conservação da *toolkit* genética e o poder dos sistemas-modelo para as descobertas biomédicas.

Câncer como uma doença do desenvolvimento

Em animais de vida longa, tais como nós mesmos e outros mamíferos, o desenvolvimento não cessa ao nascimento ou ao final da adolescência. Tecidos e diversos tipos celulares estão sendo constantemente renovados. A manutenção da função de muitos órgãos depende do crescimento controlado e da diferenciação de células que substituem aquelas que são descamadas ou de outro modo morrem. A manutenção de tecidos e órgãos em geral é controlada por vias de sinalização. Mutações hereditárias ou espontâneas nos genes que codificam os componentes dessas vias podem romper a organização tecidual e contribuir para a perda do controle da proliferação celular. Tendo em vista que a proliferação celular descontrolada é uma característica do câncer, a formação de cânceres pode ser uma consequência. O câncer, então, é uma doença do desenvolvimento, um produto dos processos normais de desenvolvimento que deram errado.

Alguns dos genes associados aos tipos de cânceres humanos são membros compartilhados da *toolkit* dos animais. Por exemplo, o gene *patched* codifica um receptor para as proteínas de sinalização Hedgehog. Além de causar distúrbios hereditários do desenvolvimento, tais como polidactilia e holoprosencefalia, mutações no gene *patched* humano estão associadas à formação de uma diversidade de cânceres. Aproximadamente 30 a 40% dos pacientes com um distúrbio genético dominante denominado *síndrome do nevo basocelular* (BCNS) apresentam mutações *patched*. Essas pessoas são fortemente predispostas ao desenvolvimento de um tipo de câncer de pele denominado carcinoma basocelular. Elas também apresentam um grande aumento da incidência de meduloblastoma, um tipo mortal de tumor cerebral. Uma crescente lista de cânceres atualmente está associada a rupturas de vias de transdução de sinal — vias que foram elucidadas pela primeira vez por triagens genéticas sistemáticas iniciais de padrões mutantes nas moscas-das-frutas (Tabela 13.2).

As descobertas de ligações entre mutações dos genes de vias de transdução de

sinal e o câncer humano facilitaram muito o estudo da biologia do câncer e o desenvolvimento de novas terapias. Por exemplo, aproximadamente 30% dos camundongos heterozigotos para uma mutação específica no gene *patched* desenvolvem meduloblastoma. Portanto, esses camundongos atuam como um excelente modelo em relação à biologia da doença humana e uma plataforma de testes para terapias. Muitos dos agentes anticâncer mais recentes empregados atualmente são, de fato, direcionados aos componentes das vias de transdução de sinal que estão comprometidos em determinados tipos de tumores.

Tabela 13.2 Alguns genes *toolkit* que apresentam papéis no câncer.

	Gene de mosca	**Gene de mamífero**	**Tipo de câncer**
Componentes de vias de sinalização			
Wingless	*armadillo*	β-*catenin*	Cólon e pele
	D.TCF	*TLF*	Cólon
Hedgehog	*cubitus interruptus*	*Gli1*	Carcinoma basocelular
	patched	*patched*	Carcinoma basocelular, meduloblastoma
	smoothened	*smoothened*	Carcinoma

			basocelular
Notch	*Notch*	*hNotch1*	Leucemia, linfoma
Receptor EGF	*torpedo*	*C-erbB-2*	Mama e cólon
Decapentaplégico/TFG-β	*Medea*	*DPC4*	Pâncreas e cólon
Toll	*dorsal*	*NF-κB*	Linfoma
Outro	*extradenticle*	*Pbx1*	Leucemia aguda de célula pré-B

É justo dizer que até mesmo os pesquisadores mais otimistas e previdentes não esperavam que a descoberta da *toolkit* genética para formar uma mosca viesse a apresentar efeitos de longo alcance sobre a compreensão do desenvolvimento e das doenças em seres humanos. Mas os grandes dividendos inesperados são familiares na história recente da pesquisa genética básica. O advento de medicamentos geneticamente modificados, anticorpos monoclonais para diagnóstico e terapia e testes de DNA forense tiveram, todos, origens semelhantes em investigações aparentemente não relacionadas.

RESUMO

No Capítulo 11, mencionamos o gracejo de Jacques Monod e François Jacob de que “qualquer coisa que se observe ser verdadeira para a *E. coli* também deve ser verdadeira para os elefantes”.[2] Agora que vimos os processos reguladores que produzem vermes, moscas, camundongos e elefantes, nós diríamos que eles

estavam certos? Se Monod e Jacob estavam se referindo ao princípio de que a transcrição gênica é controlada por proteínas reguladoras sequência-específicas, vimos que o repressor Lac bacteriano e as proteínas Hox das moscas de fato atuam de modo semelhante. Além disso, suas proteínas de ligação ao DNA apresentam o mesmo tipo de motivo. As percepções fundamentais que Jacob e Monod apresentaram a respeito do papel central do controle da transcrição gênica na fisiologia bacteriana e que esperavam se aplicar à diferenciação celular e ao desenvolvimento em organismos multicelulares complexos foram corroboradas em muitos aspectos no controle genético do desenvolvimento animal.

Muitas características em eucariotos unicelulares e multicelulares, entretanto, não são observadas em bactérias e seus vírus. Geneticistas e biólogos moleculares descobriram as funções dos íntrons, a recomposição do RNA, elementos reguladores de ação cis múltiplos e distantes, a cromatina, a recomposição alternativa e, mais recentemente, os miRNA. Ainda assim, central para o controle genético do desenvolvimento é o controle da expressão gênica diferencial.

Este capítulo apresentou uma visão geral sobre a lógica e os mecanismos para o controle da expressão gênica e do desenvolvimento em moscas-das-frutas e algumas outras espécies-modelo. Nós nos concentramos nos genes *toolkit* para os processos de desenvolvimento animal e nos mecanismos que controlam a organização de características importantes do plano corporal — o estabelecimento dos eixos corporais, a segmentação e a identidade dos segmentos. Embora tenhamos explorado apenas um número modesto de mecanismos reguladores de modo aprofundado e apenas algumas poucas espécies, semelhanças na lógica e nos mecanismos reguladores nos possibilitam identificar alguns temas gerais a respeito do controle genético do desenvolvimento.

1. *Apesar das vastas diferenças no aspecto e na anatomia, os animais apresentam em comum uma* toolkit *de genes que regulam o desenvolvimento.* Essa caixa de ferramentas é uma pequena fração de todos os genes no genoma e a maior parte dos genes dessa *toolkit* controla fatores

de transcrição e componentes das vias de transdução de sinal. Genes *toolkit* individuais apresentam múltiplas funções e afetam o desenvolvimento de diferentes estruturas em diferentes estágios.

2. *O desenvolvimento do embrião em crescimento e de suas partes corporais ocorre em uma progressão ordenada espacial e temporalmente.* São estabelecidos domínios dentro do embrião por meio da expressão dos genes *toolkit* que marcam subdivisões progressivamente mais finas ao longo de ambos os eixos embrionários.

3. *Os padrões de expressão gênica espacialmente restritos são produtos da regulação combinatória.* Cada padrão de expressão gênica apresenta uma base causal precedente. Novos padrões são gerados por meio de estímulos combinados de padrões precedentes. Nos exemplos apresentados neste capítulo, o posicionamento das listras de regra dos pares e a restrição da expressão de genes reguladores de apêndices a segmentos individuais requerem a integração de diversos estímulos reguladores positivos e negativos por parte de elementos reguladores de ação cis.

A regulação pós-transcricional no nível do RNA adiciona outra camada de especificidade ao controle da expressão gênica. A recomposição alternativa do RNA e o controle da tradução por proteínas e miRNA também contribuem para o controle espacial e temporal da expressão de genes *toolkit*.

O controle combinatório é a chave para a *especificidade* e para a *diversidade* da expressão gênica, bem como para a função de genes *toolkit*. Em relação à especificidade, mecanismos combinatórios proporcionam os meios para localizar a expressão gênica em populações celulares discretas por meio da utilização de estímulos que não são específicas do tipo celular ou tecido. As ações de proteínas toolkit, portanto, podem ser razoavelmente específicas em diferentes contextos. Em relação à diversidade, mecanismos combinatórios proporcionam os meios para gerar uma diversidade virtualmente ilimitada de padrões de expressão gênica.

4. *A modularidade dos elementos reguladores de ação cis possibilita o controle espacial e temporal independente da expressão e da função de genes toolkit.* Assim como os operadores e os elementos UAS dos

procariotos e eucariotos simples atuam como interruptores no controle fisiológico da expressão gênica, os elementos reguladores de ação cis de genes *toolkit* atuam como interruptores no controle da expressão gênica no desenvolvimento. A característica de distinção de genes *toolkit* é a presença típica de diversos elementos reguladores de ação cis independentes que regulam a expressão gênica em diferentes domínios espaciais e em diferentes estágios do desenvolvimento. A regulação espacial e temporal independente da expressão gênica possibilita que os genes *toolkit* individual apresentem funções diferentes, porém específicas, em contextos distintos. Assim, não é adequado ou preciso descrever a função de um determinado gene *toolkit* somente em relação à proteína (ou miRNA) que ele codifica, tendo em vista que a função do produto gênico quase sempre depende do contexto no qual ele é expresso.

TERMOS-CHAVE

complexo gênico
estrutura serialmente reiterada
gene de efeito materno
gene de manutenção
gene de polaridade segmentar
gene de regra dos pares
gene *gap*
gene *Hox*
homeoboxe
homeodomínio
informação posicional
toolkit genética
zigoto

PROBLEMAS RESOLVIDOS

Problema resolvido 1. O gene *Bicoid* (*bcd*) é um gene de efeito materno

necessário para o desenvolvimento da região anterior de *Drosophila*. A mãe heterozigota em relação a uma deleção *bcd* apresenta apenas uma cópia do gene *bcd*. Com a utilização de elementos *P* para inserir cópias do gene bcd^+ clonado no genoma por meio de transformação, é possível produzir mães com cópias extras do gene. O embrião inicial de *Drosophila* desenvolve uma endentação denominada sulco cefálico, que é mais ou menos perpendicular ao eixo corporal longitudinal anteroposterior (A-P). Na progênie de mães com apenas uma única cópia de bcd^+, esse sulco está muito próximo da ponta anterior, localizado a uma posição a um sexto da distância desde a ponta anterior até a posterior. Na progênie de diploides do tipo selvagem padrão (que apresentam duas cópias de bcd^+), o sulco cefálico surge mais posteriormente, em uma posição a um quinto da distância desde a ponta anterior até a posterior do embrião. Na progênie de mães com três cópias de bcd^+, ele é ainda mais posterior. Na medida em que doses adicionais do gene são adicionadas, o sulco cefálico se move mais e mais posteriormente, até que, na progênie de mães com seis cópias de bcd^+, ele está no ponto médio ao longo do eixo A-P do embrião. Explique o efeito da dosagem gênica de bcd^+ sobre a formação do sulco cefálico à luz da contribuição que *bcd* proporciona para a formação do padrão A-P.

Solução

A determinação das partes anteroposteriores do embrião é regulada por um gradiente de concentração da proteína Bicoid. O sulco se desenvolve a uma concentração crítica de *bcd*. Na medida em que a dosagem gênica de bcd^+ (e, portanto, a concentração da proteína Bicoid) diminui, o sulco é deslocado anteriormente; na medida em que a dosagem gênica aumenta, o sulco é deslocado posteriormente.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 13.2, o transplante de determinadas regiões do tecido embrionário induz o desenvolvimento de estruturas em novos locais. Como

são denominadas essas regiões especiais e quais são as substâncias que se propõe que elas produzam?

2. Na Figura 13.4, estão ilustrados dois métodos diferentes para a visualização da expressão gênica em animais em desenvolvimento. Qual método possibilitaria detectar onde, dentro de uma célula, uma proteína está localizada?
3. A Figura 13.6 ilustra a expressão da proteína Hox Ultrabithorax (Ubx) nos apêndices das asas em desenvolvimento. Qual é a relação entre o local onde a proteína é expressa e o fenótipo que resulta da perda de sua expressão (demonstrada na Figura 13.1)?
4. Na Figura 13.7, qual é a evidência de que os genes *Hox* de vertebrados regulam a identidade de estruturas serialmente repetidas?
5. Conforme demonstrado na Figura 13.10, qual é a distinção fundamental entre um gene de regra dos pares e um gene de polaridade do segmento?
6. Na Tabela 13.1, qual é a função mais comum das proteínas que contribuem para a formação do padrão? Por que esse é o caso?
7. Na Figura 13.15, qual proteína gap regula o limite posterior de *eve* da faixa 2? Descreva como ela realiza isso em termos moleculares.
8. Conforme demonstrado na Figura 13.17, quantos fatores de transcrição diferentes regulam onde o gene *Distal-less* (*Dll*) será expresso?
9. Conforme demonstrado na Figura 13.21, o gene *Sonic hedgehog* é expresso em muitos locais em uma galinha em desenvolvimento. A proteína Sonic hedgehog expressa em cada tecido é idêntica? Em caso afirmativo, como os tecidos se desenvolvem em diferentes estruturas? Em caso negativo, como as diferentes proteínas Sonic hedgehog são produzidas?

PROBLEMAS BÁSICOS

10. *Engrailed*, *even-skipped*, *hunchback* e *Antennapedia*. Para um geneticista de *Drosophila*, o que são eles? Como eles diferem?

11. Descreva o padrão de expressão do gene *eve* de *Drosophila* no embrião inicial.

12. Contraste a função dos genes homeóticos com aquela dos genes de regra dos pares.

13. Quando um embrião é homozigoto mutante em relação ao gene *gap Kr*, a quarta e a quinta listras do gene de regra dos pares *ftz* (contadas a partir da extremidade anterior) não se formam normalmente. Quando o gene *gap kni* é mutante, a quinta e a sexta listras de *ftz* não se formam normalmente. Explique esses resultados em relação ao modo como o número de segmentos é estabelecido no embrião.

14. Alguns dos genes *Hox* de mamíferos demonstraram ser mais semelhantes a um dos genes *Hox* de insetos do que a outros. Descreva uma abordagem experimental que possibilitaria a você demonstrar esse achado em um teste funcional em moscas vivas.

15. As três proteínas de homeodomínio Abd-B, Abd-A e Ubx são codificadas por genes do completo *Bithorax* de *Drosophila*. Em embriões do tipo selvagem, o gene *Abd-B* é expresso nos segmentos abdominais posteriores, *Abd-A* nos segmentos abdominais intermediários e *Ubx* nos segmentos abdominal anterior e torácico posterior. Quando o gene *Abd-B* é deletado, *Abd-A* é expresso em ambos os segmentos abdominal intermediário e posterior. Quando *Abd-A* é deletado, *Ubx* é expresso no tórax posterior e nos segmentos abdominal anterior e intermediário. Quando *Ubx* é deletado, os padrões de expressão do tipo selvagem de *Abd-A* e *Abd-B* são inalterados. Quando ambos *Abd-A* e *Abd-B* são deletados, *Ubx* é expresso em todos os segmentos, desde o tórax posterior até a extremidade posterior do embrião. Explique essas observações, levando em consideração o fato de que os genes *gap* controlam os padrões de expressão iniciais dos genes homeóticos.

16. Quais testes genéticos possibilitam que você diga se um gene é zigoticamente necessário ou se ele apresenta um efeito materno?

17. Ao considerar a formação dos eixos A-P e D-V em *Drosophila*, observamos que, em relação a mutações tais como *bcd*, mães mutantes homozigotas produzem uniformemente descendentes mutantes com defeitos de segmentação. Esse desfecho sempre é verdadeiro, independentemente de a progênie ser bcd^{+}/bcd ou *bcd/bcd*. Algumas outras mutações letais de efeito materno são diferentes, porque o fenótipo mutante pode ser "resgatado" por meio da introdução de um alelo do tipo selvagem do gene do pai. Em outras palavras, em relação aos referidos letais de efeito materno resgatáveis, animais mut^{+}/mut são normais, enquanto animais *mut/mut* apresentam o defeito mutante. Explique a diferença entre as mutações letais de efeito materno resgatáveis e não resgatáveis.

18. Suponha que você isola uma mutação que afeta a padronização A-P do embrião de *Drosophila*, na qual estão ausentes segmentos alternados das larvas mutantes em desenvolvimento.

a. Você consideraria que essa mutação se encontra em um gene *gap*, um gene de regra dos pares, um gene de polaridade do segmento ou um gene de identidade do segmento?

b. Você clonou um fragmento de DNA que contém quatro genes. Como você utilizaria o padrão de expressão espacial de seu mRNA em um embrião do tipo selvagem para identificar qual representa um gene candidato em relação à mutação descrita?

c. Presuma que você identificou o gene candidato. Se agora você examinar o padrão de expressão espacial do seu mRNA em um embrião que é homozigoto mutante em relação ao gene *gap Krüppel*, você espera observar um padrão de expressão normal? Explique.

19. Como o gradiente da proteína Bicoid é formado?

20. Em um embrião de uma fêmea mutante homozigota *Bicoid*, qual(is) classe(s) de expressão gênica é(são) normal(is)?

a. Genes *gap*.

b. Genes de regra dos pares.

c. Genes de polaridade do segmento.

d. Genes *Hox*.

PROBLEMAS DESAFIADORES

21.

a. O gene *eyeless* é necessário para a formação dos olhos em *Drosophila*. Ele codifica um homeodomínio. O que você preveria a respeito da função bioquímica da proteína Eyeless?

b. Onde você preveria que o gene *eyeless* seja expresso no desenvolvimento? Como você testaria a sua previsão?

c. Os genes *Small eye* e *Aniridia* de camundongos e seres humanos, respectivamente, codificam proteínas com sequências muito similares às proteínas Eyeless da mosca e elas são denominadas de acordo com seus efeitos sobre o desenvolvimento ocular. Planeje um teste para examinar se os genes de camundongos e seres humanos são funcionalmente equivalentes ao gene *eyeless* de mosca.

22. O gene *X* é expresso no cérebro, no coração e nos pulmões em desenvolvimento de camundongos. Mutações que afetam seletivamente a função do gene *X* nesses três tecidos são mapeadas em três regiões diferentes (A, B e C, respectivamente) a 5′ da região codificadora de *X*.

a. Explique a natureza dessas mutações.

b. Desenhe um mapa do *locus X* consistente com a informação precedente.

c. Como você testaria a função das regiões A, B e C?

23. Por que as mutações reguladoras no gene *Sonic hedgehog* de camundongos são dominantes e viáveis? Por que as mutações codificadoras causam defeitos mais difusos?

24. Ocorre uma mutação no gene *doublesex* de *Drosophila* que evita a ligação de Tra ao transcrito de RNA de *dsx*. Quais seriam as consequências dessa mutação para a expressão da proteína Dsx em machos? E em fêmeas?

25. Você isola uma mutação em *glp-1* de *C. elegans* e descobre que a região do DNA que codifica o controle espacial (SCR) foi deletada. Qual será o padrão de expressão da proteína GLP-1 em um embrião de quatro células

nos heterozigotos mutantes? E em homozigotos mutantes?

26. Avalie a validade da observação de Monod e Jacob de que “qualquer coisa que se observe ser verdadeira para a *E. coli* também deve ser verdadeira para os elefantes”.

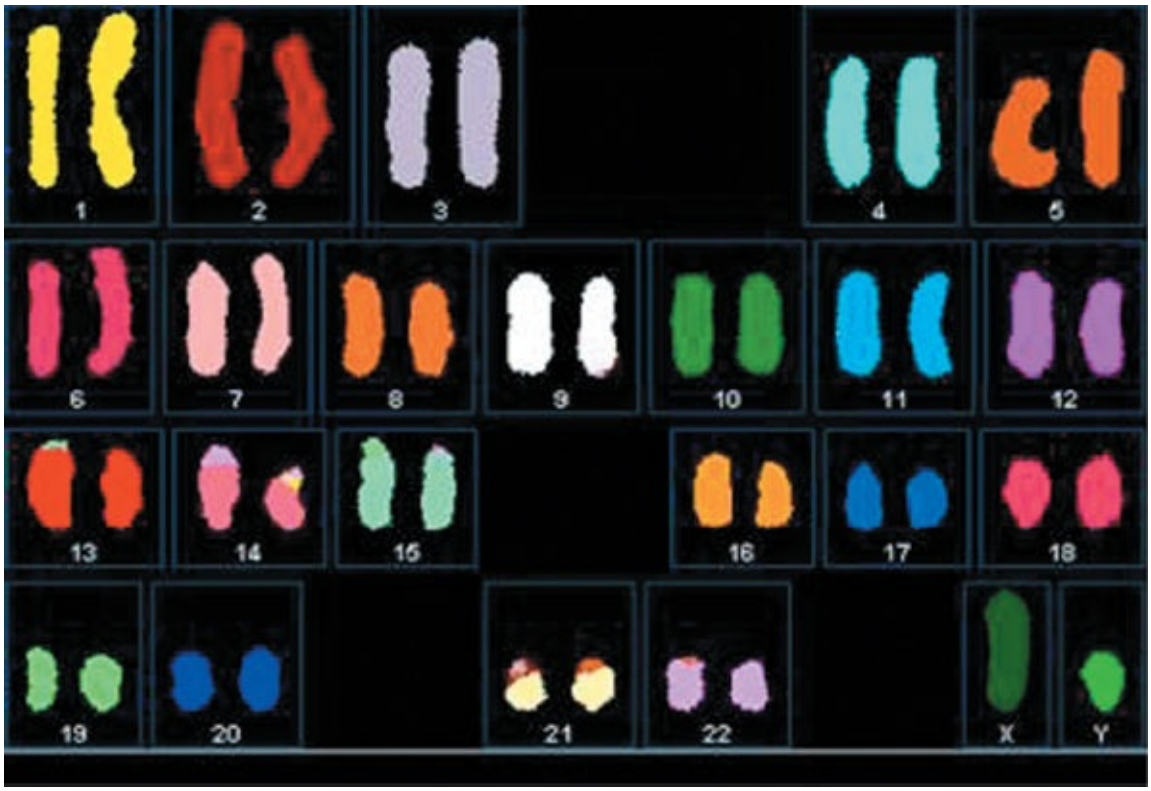
a. Compare as estruturas e os mecanismos de ação das proteínas Hox de animais e do repressor Lac. De que modos eles são semelhantes?

[1]T. H. Morgan, *Experimental Embryology.* Columbia University Press, 1927.

[2]F. Jacob e J. Monod, *Cold Spring Harbor Quant. Symp. Biol.* 26, 1963, 393.

CAPÍTULO 14

Genomas e Genômica



O genoma nuclear humano visualizado como um conjunto de DNA marcado. O DNA de cada cromossomo foi marcado com um corante que emite fluorescência em um comprimento de onda específico (produzindo uma cor específica). (Nallasivam Palanisamy, MSc., MPhil, PhD., Professor Associado de Patologia, Michigan Center for Translational Pathology, University of Michigan.)

TÓPICOS

14.1 Revolução genômica
14.2 Obtenção da sequência de um genoma
14.3 Bioinformática | Significado da sequência genômica
14.4 Estrutura do genoma humano

14.5 Genômica comparativa de seres humanos com outras espécies
14.6 Genômica comparativa e medicina de seres humanos
14.7 Genômica funcional e genética reversa

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Descrever as combinações de estratégias tipicamente necessárias para a obtenção e a montagem das sequências de DNA completas de organismos
- Listar os elementos funcionais nos genomas e explicar como eles são identificados por computador e experimentalmente
- Comparar as abordagens de genoma inteiro e subgenômicas para a medicina personalizada
- Descrever como a genômica comparativa é empregada para revelar as diferenças genéticas entre as espécies
- Explicar como a disponibilidade da sequência genômica possibilita a análise genética reversa da função dos genes.

No verão de 2009, o Dr. Alan Mayer, um pediatra do Children's Hospital of Wisconsin, em Milwaukee, escreveu para um colega a respeito do caso desconcertante e de partir o coração de um paciente seu com 4 anos de idade (Figura 14.1). Durante 2 anos, o pequeno Nicholas Volker havia realizado mais de 100 jornadas até o centro cirúrgico enquanto os médicos tentavam tratar uma doença misteriosa que estava destruindo os seus intestinos, deixando-o vulnerável a infecções perigosas, gravemente abaixo do peso e, com frequência, incapaz de comer.

Nem Mayer nem quaisquer outros médicos haviam visto uma doença como a de Nicholas; eles não conseguiam diagnosticá-la ou evitar seus danos por meio de qualquer tratamento clínico, cirúrgico ou nutricional. Era difícil tratar uma doença que ninguém conseguia identificar. Assim, o Dr. Mayer indagou ao seu colega, o Dr. Howard Jacob do Medical College of Wisconsin, "se existe algum modo de conseguirmos realizar o sequenciamento de seu genoma. Existe uma boa chance

de que o Nicholas apresente um defeito genético e é provável que seja uma nova doença. Além disso, um diagnóstico breve poderia salvar a sua vida e mostrar verdadeiramente a medicina genômica personalizada".[1]

O Dr. Jacob sabia que seria um tiro no escuro. Encontrar uma única mutação responsável por uma doença requereria peneirar milhares de variações no DNA de Nicholas. Uma decisão-chave era estreitar a busca para apenas as sequências de éxons no DNA de Nicholas. A justificativa era de que, se a mutação causal fosse uma alteração na codificação proteica, ela poderia ser identificada por meio do sequenciamento de todos os éxons ou do **exoma** de Nicholas, o que compreende pouco mais de 1% do genoma humano inteiro. Ainda assim, seria uma busca dispendiosa — o sequenciamento custaria aproximadamente US$ 75.000 com a tecnologia disponível na ocasião. Não obstante, o dinheiro foi obtido a partir de doadores, e Jacob e uma equipe de colaboradores assumiram a tarefa.

Nicholas Volker

FIGURA 14.1 O sequenciamento de todos os éxons do DNA do genoma de Nicholas Volker revelou uma mutação única responsável por sua doença debilitante, porém anteriormente não identificada. (Gary Porter/MCT/Newscom.)

Conforme Jacob esperava, eles encontraram mais de 16.000 possíveis variações candidatas no DNA de Nicholas. Estreitaram essa longa lista ao enfocar as mutações que não haviam sido identificadas anteriormente em seres humanos e que causavam substituições de aminoácidos que não eram observadas em outras espécies. Finalmente, eles identificaram uma única substituição de bases em um gene denominado *i*nibidor da *ap*optose ligado ao *X* (*XIAP*) que alterava um aminoácido na posição 203 da proteína — um aminoácido que era invariante nos correspondentes do gene *XIAP* entre mamíferos, peixes e mesmo nas mosca-das-frutas.

Felizmente, a identificação da mutação no gene *XIAP* de Nicholas sugeriu uma abordagem terapêutica. Sabia-se anteriormente que o gene *XIAP* apresenta um papel na resposta inflamatória e mutações no gene estavam associadas a um distúrbio imune muito raro, porém possivelmente fatal (embora não aos sintomas intestinais de Nicholas). Com base nesse conhecimento, os médicos de Nicholas atacaram o seu sistema imune com uma infusão de sangue de cordão umbilical de um doador compatível. Durante os diversos meses que se seguiram, a saúde de Nicholas melhorou até o ponto em que ele conseguia comer carne e outros alimentos. E ao longo dos 2 anos seguintes, Nicholas não precisou de quaisquer cirurgias intestinais adicionais.

O diagnóstico e o tratamento de Nicholas Volker ilustram os avanços dramáticos na tecnologia e o impacto da **genômica** — o estudo dos genomas em sua totalidade. A promessa há muito esperada de que a genômica modelaria a medicina clínica agora é muito mais uma realidade. O progresso tecnológico e biológico que teve início com alguns dados na década de 1990 foi impressionante. Em 1995, o genoma de 1,8 Mb (1,8 megabase) da bactéria *Haemophilus influenzae* foi o primeiro genoma de um organismo vivo a ser sequenciado. Em 1996, surgiu o genoma de 12 Mb de *Saccharomyces cerevisiae*; em 1998, o genoma de 100 Mb de *C. elegans*; em 2000, o genoma de 180 Mb de *Drosophila melanogaster*; em 2001, o primeiro rascunho do genoma humano de

3.000 Mb; e, em 2005, o primeiro rascunho de nosso parente vivo mais próximo, o chimpanzé. Essas espécies são apenas uma pequena amostra. No final de 2013, as sequências de quase 27.000 genomas bacterianos e mais de 6.600 espécies eucarióticas (incluindo fungos, plantas e animais) haviam sido decifradas.

Não é uma hipérbole dizer que a genômica revolucionou o modo como a análise genética é realizada e abriu vias de indagações que não eram concebíveis há alguns anos. A maior parte das análises genéticas que consideramos até então emprega uma abordagem direta para a análise de processos genéticos e biológicos. Ou seja, a análise tem início primeiramente com a triagem de mutantes que afetam algum fenótipo observável e a caracterização desses mutantes finalmente leva à identificação do gene e da função do DNA, do RNA e das sequências proteicas. Contrariamente, conhecer as sequências inteiras de DNA do genoma de um organismo possibilita que os geneticistas trabalhem em ambos os sentidos — diretamente do fenótipo até o gene e de modo reverso do gene até o fenótipo. Sem exceções, as sequências genômicas revelam muitos genes que não foram detectados a partir da análise mutacional clássica. Com a utilização da assim denominada genética reversa, atualmente os geneticistas conseguem estudar sistematicamente os papéis dos referidos genes não identificados anteriormente. Além disso, a ausência de estudo genético clássico prévio deixou de ser um impedimento para a investigação genética dos organismos. As fronteiras da análise experimental estão crescendo muito além dos limites do número muito modesto de organismos-modelo há muito explorados.

As análises de genomas inteiros atualmente contribuem para todos os aspectos das pesquisas biológicas. Na genética humana, a genômica está proporcionando novos modos para localizar genes que contribuem para muitas doenças genéticas, como a de Nicholas, que anteriormente haviam escapado dos investigadores. Logo chegará o dia em que a sequência do genoma de uma pessoa será uma parte padrão de seu registro médico. A disponibilidade das sequências genômicas de organismos-modelo estudados há muito tempo e seus parentes acelerou dramaticamente a identificação dos genes, a análise da função gênica e a caracterização de elementos não codificadores do genoma. Novas tecnologias para a análise global do papel fisiológico de todos os produtos gênicos do

genoma estão direcionando o desenvolvimento do novo campo denominado *biologia de sistemas*. A partir de uma perspectiva evolutiva, a genômica proporciona uma visão detalhada sobre como os genomas e os organismos divergiram e se adaptaram ao longo do tempo geológico.

A sequência de DNA do genoma é o ponto de início para todo um novo conjunto de análises direcionadas à compreensão da estrutura, da função e da evolução do genoma e de seus componentes. Neste capítulo, enfocaremos três aspectos importantes da análise genômica:

- *Bioinformática*, a análise do conteúdo das informações de genomas inteiros. Essas informações incluem os números e os tipos de genes e produtos gênicos, bem como a localização, o número e os tipos de sítios de ligação no DNA e no RNA que possibilitam aos produtos funcionais serem sintetizados na ocasião e no local corretos
- *Genômica comparativa*, que considera os genomas de espécies relacionadas de modo próximo e distante a partir de uma percepção evolutiva
- *Genômica funcional*, a utilização de uma diversidade de métodos em expansão, incluindo genética reversa, para compreender a função dos genes e das proteínas nos processos biológicos.

14.1 Revolução genômica

Após o desenvolvimento da tecnologia do DNA recombinante na década de 1970, laboratórios de pesquisa passaram a fazer a clonagem e o sequenciamento de um gene por vez, prosseguindo apenas após terem primeiramente encontrado algo interessante a respeito daquele gene a partir de uma análise mutacional clássica. As etapas desde a elaboração do mapa genético clássico de um *locus* para isolar o DNA codificador de um gene (*clonagem*) até a determinação de sua sequência com frequência eram numerosas e tomavam tempo. Na década de 1980, alguns cientistas perceberam que uma grande equipe de pesquisadores, em um esforço conjunto, poderia clonar e sequenciar o genoma *inteiro* de um organismo selecionado. Tais **projetos genoma** tornariam, então, os clones e a sequência

recursos publicamente disponíveis. Um atrativo da disponibilização desses recursos é que, quando os pesquisadores se interessam por um gene de uma espécie cujo genoma foi sequenciado, eles apenas precisam descobrir onde esse gene está localizado no mapa do genoma para conseguir enfocar sua sequência e possivelmente sua função. Assim, um gene pode ser caracterizado muito mais rapidamente do que por meio da sua clonagem e do seu sequenciamento a partir do zero, algo que na época poderia demorar muitos anos para ser realizado. Essa abordagem mais rápida atualmente é uma realidade em relação a todos os organismos-modelo.

De modo similar, o Projeto Genoma Humano teve por objetivo revolucionar o campo da genética humana. A disponibilidade de sequências do genoma humano e a capacidade de sequenciar os genomas de pacientes e seus parentes auxiliaram muito na identificação dos genes causadores de doenças. De modo semelhante, a capacidade de determinar as sequências gênicas em tecidos normais e enfermos (p. ex., cânceres) tem sido um grande catalisador para a compreensão dos processos de doença e apontou o caminho para novas terapias.

De uma perspectiva mais ampla, os projetos genoma tinham o atrativo de proporcionar um vislumbre dos princípios pelos quais os genomas são construídos. O genoma humano contém 3 bilhões de pares de bases de DNA. A sequência inteira decifrada deu origem a questões como: quantos genes ele contém? Como eles estão distribuídos e por quê? Que fração do genoma é a sequência codificadora? Que fração é a sequência reguladora? Quanto do nosso genoma é similar ou diferente do de outros animais? Embora estejamos convictos de que podemos compreender um único gene de interesse, o principal desafio da genômica atualmente é a literatura genômica: como lemos o depósito de informações codificadas na sequência de genomas completos?

As técnicas básicas necessárias para o sequenciamento de genomas inteiros já estavam disponíveis na década de 1980 (ver Capítulo 10). Mas a escala necessária para sequenciar um genoma complexo estava, assim como um projeto de engenharia, muito além da capacidade de pesquisa da comunidade científica daquela época. A genômica no final da década de 1980 e na década de 1990 evoluiu a partir de grandes centros de pesquisas que conseguiram integrar essas

tecnologias elementares em uma linha de produção de nível industrial. Esses centros desenvolveram robótica e automação para realizar as milhares de etapas de clonagem e milhões de reações de sequenciamento necessárias para montar a sequência de um organismo complexo. Tão importante quanto isso, avanços na tecnologia da informação auxiliaram na análise dos dados resultantes.

Os primeiros sucessos no sequenciamento do genoma compensaram as ondas de inovação que levaram às tecnologias de sequenciamento mais rápidas e muito menos dispendiosas. Atualmente, máquinas individuais podem produzir tantas sequências em 1 dia quanto os centros costumavam realizar em meses. Atualmente, novas tecnologias podem obter mais de 100 bilhões de sequências de bases em 1 dia de trabalho de um único instrumento. Esse número representa um aumento de aproximadamente *100.000 vezes* na produtividade em relação aos instrumentos mais iniciais utilizados para obter a primeira sequência do genoma humano.

A genômica, auxiliada pelo crescimento explosivo na tecnologia da informação, encorajou os pesquisadores a desenvolverem modos de experimentação no genoma como um todo, em vez de simplesmente um gene por vez. A genômica também demonstrou o valor da coleta antecipada de conjuntos de dados em grande escala, de modo que eles possam ser utilizados posteriormente para abordar problemas de pesquisa específicos. Nas últimas seções deste capítulo, exploraremos alguns modos pelos quais atualmente a genômica direciona pesquisas genéticas básicas e aplicadas. Nos capítulos subsequentes, veremos como a genômica está catalisando os avanços na compreensão da dinâmica da mutação, da recombinação e da evolução.

> CONCEITO-CHAVE A caracterização de genomas inteiros é fundamental para a compreensão de todo o conjunto de informação genética subjacente à fisiologia e ao desenvolvimento dos organismos vivos e para a descoberta de novos genes, tais como aqueles que apresentam papéis em doenças genéticas humanas.

14.2 Obtenção da sequência de um genoma

Quando as pessoas encontram novos territórios, uma de suas primeiras atividades é criar um mapa. Essa prática tem sido frequente para exploradores, geógrafos, oceanógrafos e astrônomos, sendo também utilizada pelos geneticistas. Os geneticistas utilizam muitos tipos de mapas para explorar o terreno de um genoma. Exemplos são os mapas de ligação com base nos padrões de herança de alelos gênicos e os mapas citogenéticos com base na localização de características microscopicamente visíveis, tais como pontos de quebra de rearranjos.

O mapa de mais alta resolução é a sequência completa do DNA do genoma — ou seja, a sequência completa de nucleotídios A, T, C e G de cada dupla-hélice no genoma. Tendo em vista que a obtenção da sequência completa de um genoma é uma realização tão maciça de um tipo não observado anteriormente na biologia, novas estratégias devem ser utilizadas, todas com base na automação.

Transformação das leituras de sequências em uma sequência montada

Você provavelmente viu um ato de mágica no qual o mágico corta uma página de jornal em muitas partes, mistura-as em seu chapéu, diz algumas palavras mágicas e *voilà!*, reaparece uma página de jornal intacta. Basicamente, é assim que as sequências genômicas são obtidas. A abordagem é (1) quebrar as moléculas de DNA de um genoma em milhares a milhões de pequenos segmentos sobrepostos mais ou menos aleatórios; (2) ler a sequência de cada pequeno segmento; (3) encontrar, de modo computadorizado, a sobreposição entre os pequenos segmentos na qual suas sequências são idênticas; e (4) continuar a realizar a sobreposição de peças cada vez maiores até que todos os pequenos segmentos estejam ligados (Figura 14.2). Nesse ponto, a sequência de um genoma está montada.

Por que esse processo requer automação? Para compreender o motivo, consideraremos o genoma humano, que contém aproximadamente 3×10^9 pb de DNA ou 3 bilhões de pares de bases (3 gigapares de bases = 3 Gpb). Suponha que possamos purificar o DNA intacto de cada um dos 24 cromossomos humanos (X, Y e os 22 autossomos), posicionar em separado cada uma dessas 24 amostras

de DNA em uma máquina de sequenciamento e ler as suas sequências diretamente a partir de um telômero até o outro. A obtenção de uma sequência completa seria totalmente direta, como a leitura de um livro com 24 capítulos — apesar de ser um livro muito, muito longo, com 3 bilhões de caracteres (aproximadamente o comprimento de 3.000 romances). Infelizmente, não existe tal máquina de sequenciamento.

Em vez disso, o sequenciamento automatizado é a atual tecnologia de ponta no sequenciamento do DNA. Inicialmente baseado no pioneiro método de Sanger de terminação de cadeia didesóxi (ver Capítulo 10), o sequenciamento automatizado atualmente emprega uma diversidade de métodos químicos e de detecção óptica. Os métodos atualmente disponíveis variam no comprimento da sequência de DNA obtida, nas bases determinadas por segundo e na precisão bruta. Para os projetos de sequenciamento em grande escala que buscam analisar grandes genomas individuais ou os genomas de muitos indivíduos ou espécies diferentes, a escolha de um método requer o equilíbrio entre a velocidade, o custo e a precisão.

Reações de sequenciamento individuais (denominadas *leituras de sequenciamento*) fornecem sequências de letras que, dependendo da técnica de sequenciamento empregada, variam em média desde aproximadamente 100 a 5.000 bases de comprimento. Tais comprimentos são pequenos em comparação ao DNA de um único cromossomo. Por exemplo, uma leitura individual de 300 bases é apenas 0,0001% do mais longo cromossomo humano (aproximadamente 3×10^8 pb de DNA) e apenas aproximadamente 0,00001% do genoma humano inteiro. Portanto, um grande desafio enfrentado pelo projeto genoma é a **montagem da sequência** — ou seja, a junção de todas as leituras individuais em uma **sequência consenso**, uma sequência em relação à qual existe consenso (ou concordância) de que essa é uma representação autêntica da sequência para cada uma das moléculas de DNA naquele genoma.

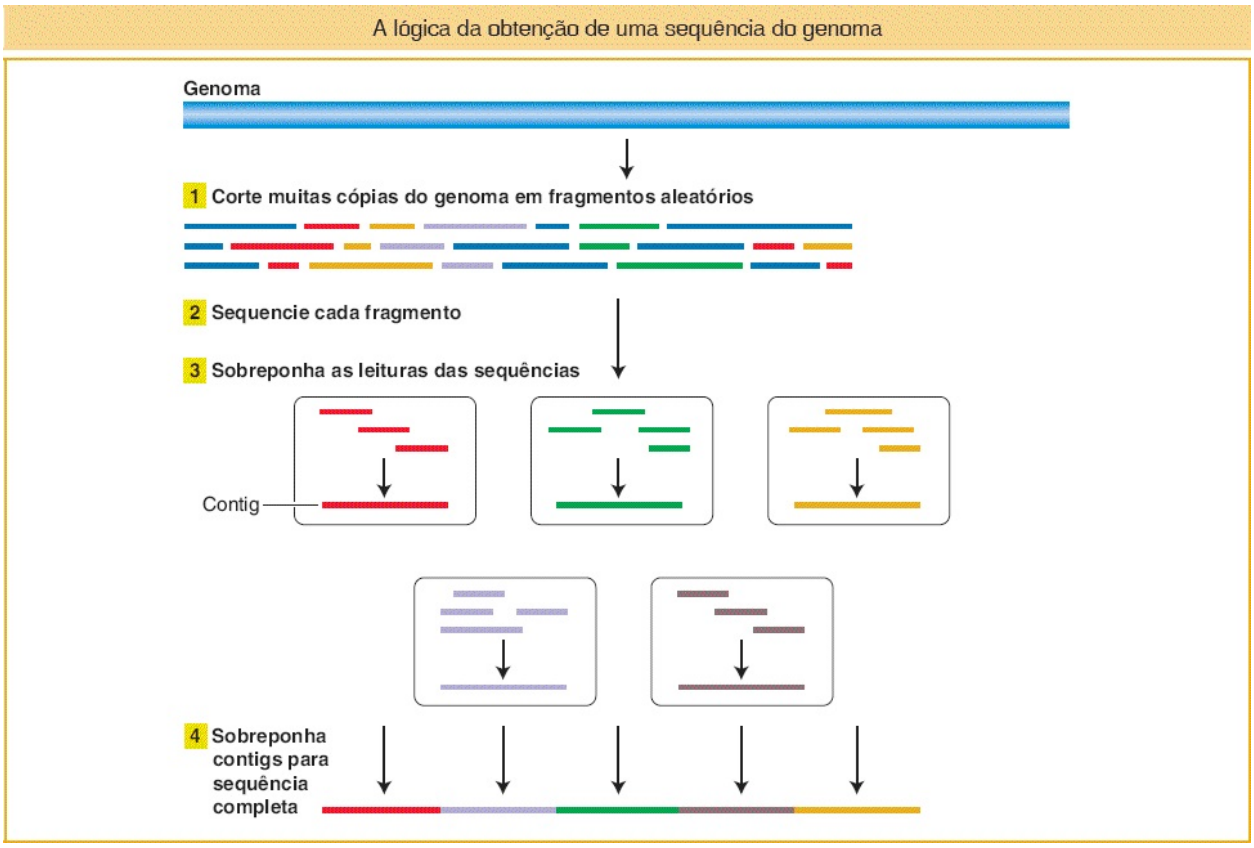


FIGURA 14.2 Para obter uma sequência genômica, múltiplas cópias do genoma são cortadas em pequenos pedaços que são sequenciados. As leituras das sequências resultantes são sobrepostas por meio da correspondência de sequências idênticas em fragmentos diferentes, até que seja produzida uma sequência consenso de cada dupla-hélice do DNA no genoma.

Vejamos esses números de um modo razoavelmente diferente para compreender a escala do problema. Assim como qualquer observação experimental, as máquinas de sequenciamento automatizado nem sempre fornecem leituras de sequência perfeitamente precisas. De fato, tecnologias de sequenciamento mais novas e de mais alta produtividade geram uma frequência *maior* de erros do que os métodos mais antigos; a taxa de erro pode variar de menos de 1% até tanto quanto 10%, dependendo da tecnologia. Portanto, para assegurar a precisão, os projetos genoma convencionalmente obtêm muitas leituras de sequência independentes de cada par de bases em um genoma. A cobertura de muitas vezes assegura que os erros ao acaso nas leituras não forneçam uma reconstrução falsa da sequência consenso.

Considerando uma leitura de sequência média de aproximadamente 100 bases de DNA e um genoma humano de 3 bilhões de pares de bases, são necessários 300 milhões de leituras independentes para proporcionar uma cobertura média de 10 vezes de cada par de bases. Entretanto, nem todas as sequências são representadas igualmente e, assim, o número de leituras necessárias é maior. A quantidade de informação a ser rastreada é enorme. Portanto, o sequenciamento do genoma necessitou de muitos avanços na automação e na tecnologia da informação.

Quais são os objetivos do sequenciamento de um genoma? Primeiramente, nos esforçamos para produzir uma sequência consenso que seja uma representação verdadeira e precisa do genoma, iniciando com um organismo individual ou uma linhagem-padrão a partir dos quais o DNA tenha sido obtido. Essa sequência atuará então como uma sequência de referência em relação à espécie. Atualmente, sabemos que existem muitas diferenças na sequência de DNA entre diferentes indivíduos de uma espécie e até mesmo entre os genomas de origem materna e paterna em um único indivíduo diploide. Portanto, nenhuma sequência do genoma representa verdadeiramente o genoma de toda a espécie. Não obstante, a sequência do genoma atua como um padrão ou uma referência à qual outras sequências podem ser comparadas, e pode ser analisada para determinar a informação codificada no DNA, tal como o inventário dos RNA e polipeptídios codificados.

Assim como os manuscritos, as sequências genômicas podem variar desde a qualidade *rascunho* (o esboço geral está presente, mas existem erros tipográficos, erros gramaticais, lacunas, seções que necessitam de rearranjo e assim por diante), até a qualidade *final* (uma taxa muito baixa de erros tipográficos, algumas seções ausentes, mas tudo o que atualmente é possível foi feito para preencher essas seções), até a verdadeiramente *completa* (nenhum erro tipográfico, cada par de bases absolutamente correto de telômero a telômero). Nas seções a seguir, examinaremos a estratégia e alguns métodos para a produção de montagens de sequências genômicas rascunho e finalizadas. Também encontraremos algumas das características dos genomas que desafiam os projetos de sequenciamento genômico.

Sequenciamento do genoma inteiro

A atual estratégia geral para a obtenção e a montagem da sequência de um genoma é denominada *sequenciamento shotgun de genoma inteiro* (*WGS*; do inglês, *whole-genome shotgun*). Essa abordagem tem por base a determinação da sequência de muitos segmentos de DNA genômico que foram gerados por meio da quebra de DNA de longos cromossomos em muitos segmentos curtos. Duas abordagens do sequenciamento *shotgun* de genoma inteiro são responsáveis pela maior parte das sequências genômicas obtidas até agora. As diferenças fundamentais entre elas estão no modo como os segmentos curtos de DNA são obtidos e preparados para o sequenciamento e na química empregada no sequenciamento. O primeiro método utilizado para sequenciar o primeiro genoma humano baseou-se na clonagem do DNA em células microbianas e empregou a técnica de sequenciamento didesóxi de Sanger. Faremos referência a essa abordagem como "WGS tradicional". No segundo grupo, os métodos em geral são livres de células, empregam novas técnicas para o sequenciamento e são desenhados para processamento muito alto (que se refere ao número de leituras por máquina por unidade de tempo). Faremos referência a esse grupo de métodos como "WGS de última geração".

WGS tradicional

A abordagem WGS tradicional tem início com a construção de **bibliotecas genômicas**, que são coleções desses segmentos de DNA curtos, representando o genoma inteiro. Os segmentos de DNA curtos em tal biblioteca foram inseridos em um de uma diversidade de tipos de cromossomos *acessórios* (elementos não essenciais, tais como plasmídios, vírus bacterianos modificados ou cromossomos artificiais) e propagados em microrganismos, normalmente bactérias ou leveduras. Esses cromossomos acessórios que carregam insertos de DNA são denominados **vetores**.

Para gerar uma biblioteca genômica, um pesquisador primeiramente utiliza *enzimas de restrição*, que clivam o DNA em sequências específicas, para cortar o DNA genômico purificado. Algumas enzimas cortam o DNA em muitos locais,

enquanto outras o cortam em menos locais. Assim, o pesquisador consegue controlar se o DNA é cortado, em média, em pedaços mais longos ou mais curtos. Os fragmentos resultantes apresentam filamentos únicos curtos de DNA em ambas as extremidades. Cada fragmento em seguida é unido à molécula de DNA do cromossomo acessório, que também foi cortado com uma enzima de restrição e que apresenta extremidades que são complementares àquelas dos fragmentos genômicos. Para que o genoma inteiro seja representado, múltiplas cópias do DNA genômico são cortadas em fragmentos. Assim, são gerados milhares a milhões de diferentes moléculas recombinantes fragmento-vetor.

O resultante conjunto de moléculas de DNA recombinante é propagado em seguida, tipicamente por meio da introdução das moléculas em células bacterianas. Cada célula capta uma molécula recombinante. Em seguida, cada molécula recombinante é replicada durante o crescimento e a divisão normais de seu hospedeiro, de modo que são produzidas muitas cópias idênticas do fragmento inserido para a utilização na análise da sequência do fragmento de DNA. Tendo em vista que cada molécula recombinante é amplificada a partir de uma célula individual, cada célula é um *clone* distinto. (Mais detalhes a respeito da clonagem de DNA são fornecidos no Capítulo 10.) A biblioteca de clones resultante é denominada *biblioteca shotgun*, tendo em vista que as sequências lidas são obtidas a partir de clones selecionados aleatoriamente da biblioteca do genoma inteiro sem qualquer informação sobre onde esses clones estão mapeados no genoma.

Em seguida, os fragmentos genômicos nos clones da biblioteca *shotgun* são parcialmente sequenciados. A reação de sequenciamento deve ter início a partir de um *primer* de sequência conhecida. Tendo em vista que a sequência de um inserto clonado não é conhecida (e é o objetivo do exercício), os *primers* baseiam-se na sequência de DNA do vetor adjacente. Esses *primers* são utilizados para direcionar a reação de sequenciamento do inserto. Portanto, regiões curtas em uma das ou ambas as extremidades dos insertos genômicos podem ser sequenciadas (Figura 14.3). Após o sequenciamento, obtém-se uma grande coleção de sequências curtas aleatórias, algumas delas sobrepostas. Essas sequências são montadas em uma sequência consenso que abrange o genoma

inteiro por meio da correspondência de sequências homólogas compartilhadas por leituras de clones sobrepostos. As sequências de leituras sobrepostas são montadas em unidades denominadas **sequências contigs** (sequências que são contíguas ou que se tocam).

Sequenciamento shotgun de última geração do genoma inteiro

O objetivo do WGS de nova geração é o mesmo daquele do WGS tradicional — obter um grande número de leituras de sequências sobrepostas que possam ser montadas em *contigs*. Entretanto, as metodologias utilizadas diferem em diversos modos substanciais do WGS tradicional. Foram desenvolvidos diversos sistemas diferentes que, embora sejam diferentes na sua química de sequenciamento e no desenho de sua máquina, empregam, cada um, três estratégias que aumentaram dramaticamente a produtividade:

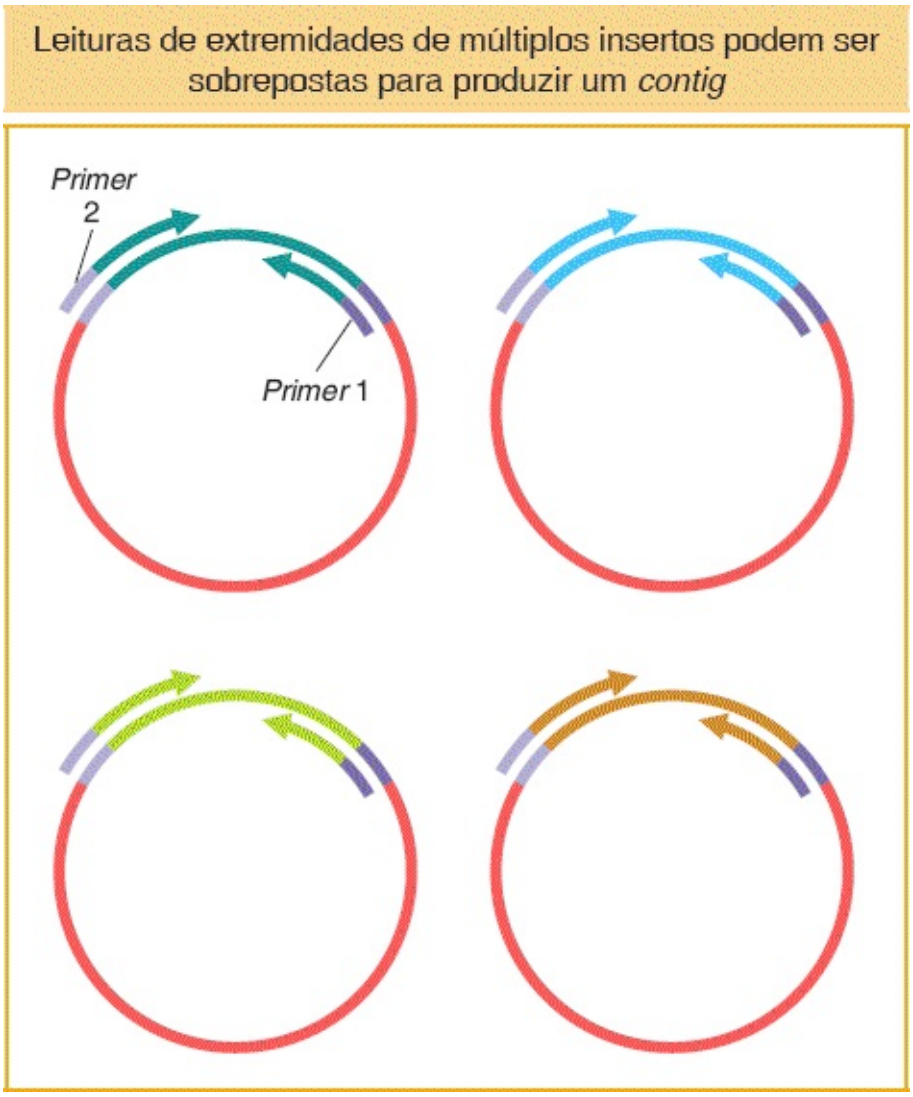


FIGURA 14.3 As leituras de sequenciamento são obtidas apenas das extremidades dos insertos clonados. O uso de sequências de dois sítios de iniciação diferentes, um em cada extremidade do vetor, torna possível o sequenciamento de até 600 pares de bases em cada extremidade do inserto genômico. Se ambas as extremidades do mesmo clone forem sequenciadas, as duas leituras de sequência resultantes são denominadas *leituras de pontas pareadas*.

1. As moléculas de DNA são preparadas para o sequenciamento em reações *livres de células*, sem clonagem em hospedeiros microbianos.
2. Milhões de fragmentos de DNA individuais são isolados e sequenciados em paralelo durante cada operação da máquina.
3. Tecnologias de manipulação de líquido, câmeras e *software* avançados tornam possível detectar os produtos de reações de sequenciamento em

volumes de reação extremamente pequenos.

Tendo em vista que o campo da tecnologia genômica está evoluindo rapidamente, não descreveremos todos os sistemas de nova geração. Entretanto, examinaremos uma abordagem amplamente utilizada que emprega todas essas características. Um dos primeiros sistemas de nova geração foi o 454, desenvolvido pela Life Sciences Corporation. Essa abordagem ilustra os ganhos que foram obtidos na produtividade e o que os referidos ganhos possibilitam que os geneticistas realizem. Pode-se considerar que a abordagem tenha três estágios:

Estágio 1. É construída uma **biblioteca de DNA molde** de moléculas de DNA unifilamentares.

Estágio 2. As moléculas de DNA na biblioteca molde são amplificadas em muitas cópias, não por meio do crescimento de colônias como nas bibliotecas genômicas tradicionais, mas por meio da utilização da reação da cadeia de polimerase (PCR; ver Capítulo 10). Primeiramente, moléculas únicas são imobilizadas em esferas individuais (*beads*). Em seguida, as moléculas são amplificadas por meio de PCR de modo que as moléculas de DNA unifilamentares permaneçam unidas aos *beads*. Assim, cada *bead* contém muitos fragmentos de DNA idênticos. Em seguida, cada *bead* é depositado individualmente dentro de poços de volume muito pequeno em um dispositivo no qual ocorrem as reações de sequenciamento (Figura 14.4).

Estágio 3. O sequenciamento de cada *bead* é realizado com a utilização de uma nova química de "sequenciamento por síntese" denominada **pirossequenciamento** (Figura 14.5). A DNA polimerase e um *primer* são adicionados aos poços para iniciar a síntese de um filamento de DNA complementar. Cada um dos quatro desoxirribonucleotídios, dATP, dGTP, dTTP e dCTP, flui por todos os poços, um por vez, em uma ordem específica. Quando é adicionado um nucleotídio que é complementar à próxima base no filamento-molde em um determinado poço, ele é incorporado e a reação libera uma molécula de pirofosfato. Duas enzimas, *sulfurilase* e *luciferase*, que também estão presentes, em seguida atuam para

converter o sinal de pirofosfato em um sinal de luz visível (ver Figura 14.5). A luz é detectada por meio de uma câmera especial. Portanto, os filamentos de DNA em crescimento que apresentam A como a primeira base após o *primer* produzirão um sinal apenas quando o dATP fluir através do poço, não quando os outros desoxinucleotídios fluírem. A reação é repetida por no mínimo 100 ciclos e os sinais de cada poço ao longo de todos os ciclos são integrados para gerar as leituras de sequência de cada poço.

Outras plataformas amplamente utilizadas, tais como os sistemas de sequenciamento Illumina e da Pacific Bioscience, também detectam a síntese de DNA, mas por diferentes meios. O sistema Illumina detecta a incorporação de dNTP individuais marcados por fluorescência, enquanto o processo da Pacific Bioscience detecta bases que estão sendo incorporadas em uma única molécula de DNA imobilizada. O método escolhido pelos investigadores depende muito da aplicação. O sistema Illumina produz uma quantidade maior de leituras mais curtas em relação ao sistema 454, enquanto o sistema da Pacific Bioscience proporciona a vantagem de leituras individuais muito mais longas do que qualquer outro sistema, mas com uma taxa de erros mais alta. A alta produtividade de cada sistema é produto do sequenciamento maciçamente em paralelo: diversas centenas de milhares a mais de 1 milhão de reações podem ocorrer simultaneamente. As máquinas de sequenciamento precedentes eram capazes de realizar apenas 384 reações de sequenciamento por operação.

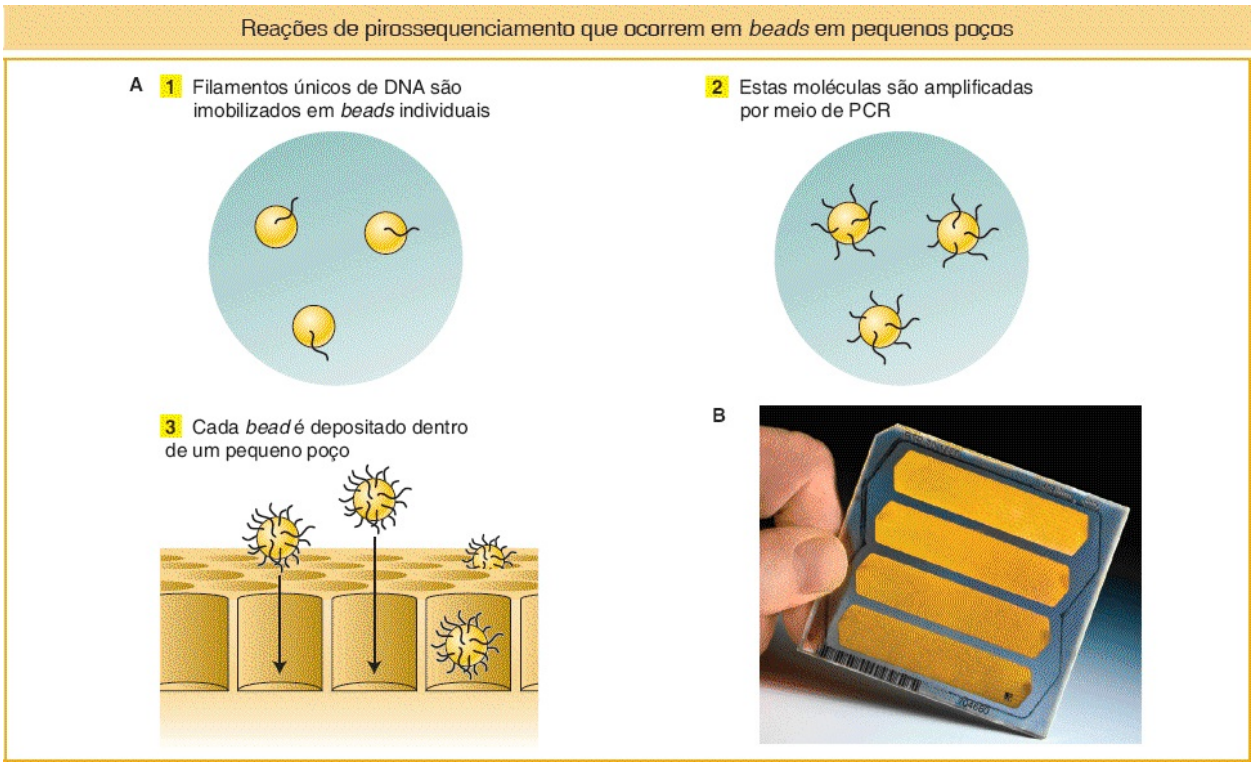


FIGURA 14.4 A. No sistema de sequenciamento 454, DNA unifilamentares são replicados em pequenos *beads* no preparo para o sequenciamento. **B.** As reações de sequenciamento de pirossequenciamento ocorrem em pequenos poços dispostos em placas. Os muitos poços em uma placa e os volumes de reação muito pequenos possibilitam o sequenciamento paralelo maciço do DNA a um custo modesto. (*B.* © 2010 The Regents of the University of California, Lawrence Berkeley National Laboratory.)

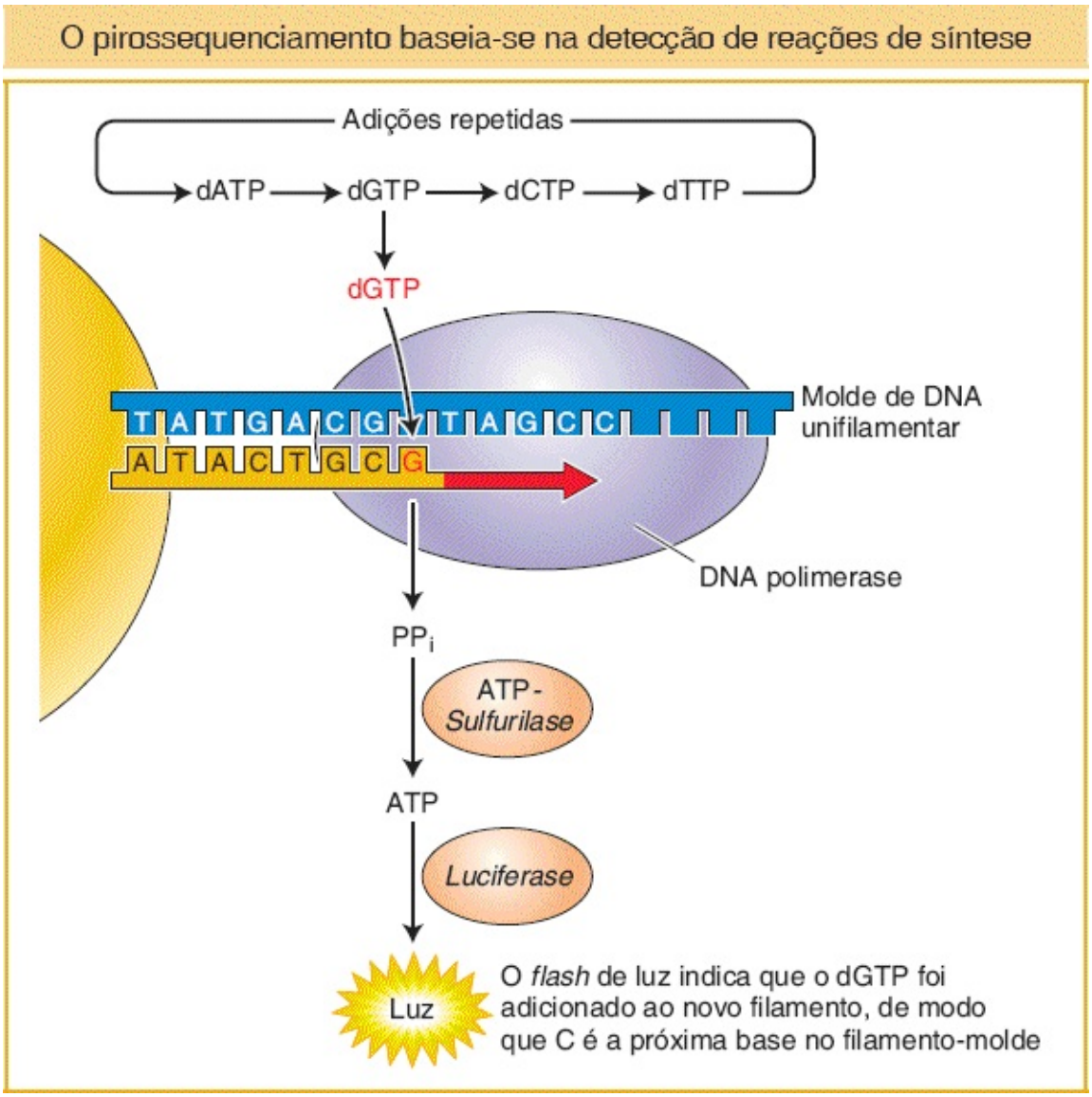


FIGURA 14.5 No processo de pirossequenciamento, nucleotídios são adicionados sequencialmente para formar o filamento complementar ao molde unifilamentar, ao qual um *primer* de sequenciamento foi pareado. As reações são realizadas na presença das enzimas DNA polimerase, sulfurilase e luciferase. Uma molécula de pirofosfato (PP_i) é liberada para cada nucleotídio incorporado ao filamento em crescimento pela DNA polimerase e é convertida em ATP pela sulfurilase. A luz visível é produzida a partir da luciferina em uma reação catalisada pela luciferase que utiliza o ATP produzido pela sulfurilase.

Montagem da sequência do genoma inteiro

Seja qual for o método de obtenção de sequência bruta utilizado, permanece o desafio de montar os contigs na sequência do genoma inteiro. A dificuldade de tal processo depende fortemente do tamanho e da complexidade do genoma.

Por exemplo, os genomas de espécies bacterianas são relativamente fáceis de montar. O DNA bacteriano é essencialmente um DNA de *cópia única*, sem sequências repetidas. Portanto, qualquer sequência de DNA lida de um genoma bacteriano terá vindo de um único lugar naquele genoma. Em virtude dessas propriedades, contigs nos genomas bacterianos com frequência podem ser montados em contigs maiores que representam a maior parte ou toda a sequência genômica de modo relativamente direto. Além disso, um genoma bacteriano típico é do tamanho de apenas alguns pares de megabases de DNA.

Em relação aos eucariotos, a montagem do genoma normalmente apresenta algumas dificuldades. Um grande obstáculo é a existência de diversas classes de sequências repetidas, algumas dispostas em *tandem* e outras dispersas. Por que elas são um problema para o sequenciamento do genoma? Resumidamente, porque uma leitura de sequenciamento de DNA repetitivo se encaixa dentro de muitos locais no rascunho do genoma. Não raramente, uma sequência repetitiva em *tandem* no total é mais longa do que o comprimento de uma leitura de sequência máxima. Nesse caso, não existe como preencher a lacuna entre sequências únicas adjacentes. Elementos repetitivos dispersos podem causar o alinhamento errôneo conjunto de leituras de diferentes cromossomos ou diferentes partes do mesmo cromossomo.

> CONCEITO-CHAVE O cenário dos cromossomos eucarióticos inclui uma diversidade de segmentos de DNA repetitivos. Esses segmentos são de difícil alinhamento como leituras de sequência.

O sequenciamento *shotgun* de genoma inteiro é particularmente bom para a produção de sequências de qualidade rascunho de genomas complexos com muitas sequências repetitivas. Como um exemplo, consideraremos o genoma da mosca-das-frutas *D. melanogaster*, que foi inicialmente sequenciado pelo método de WGS tradicional. O projeto teve início com o sequenciamento de bibliotecas de clones genômicos de diferentes tamanhos (2 kb, 10 kb, 150 kb). As leituras de sequência foram obtidas a partir de *ambas* as extremidades dos insertos de clones

genômicos e alinhadas por meio de uma lógica idêntica àquela utilizada para o sequenciamento WGS bacteriano. Por meio dessa lógica, foram identificadas as sobreposições de sequências e os clones foram posicionados em ordem, produzindo sequências contigs — sequências consenso em relação a esses trechos de cópia única do genoma. Entretanto, contrariamente à situação em bactérias, os contigs finalmente atingiram um segmento do DNA repetitivo impedindo a montagem não ambígua dos contigs em um genoma inteiro. As sequências contigs apresentavam um tamanho médio de aproximadamente 150 kb. O desafio então era como unir as milhares de tais sequências contigs em sua ordem e orientação corretas.

A solução para esse problema foi utilizar os pares de leituras de sequência de extremidades opostas dos insertos genômicos no mesmo clone — essas leituras são denominadas **leituras de pontas pareadas**. A ideia era encontrar leituras de pontas pareadas que abrangessem os intervalos entre duas sequências contigs (Figura 14.6). Em outras palavras, se uma ponta de um inserto fosse parte de um contig e a outra extremidade fosse parte de um segundo contig, então esse inserto deveria abranger o intervalo entre dois contigs e os dois contigs estariam claramente próximos um do outro. De fato, tendo em vista que o tamanho de cada clone era conhecido (ou seja, ele tinha origem em uma biblioteca que continha insertos genômicos de tamanho uniforme, fosse a biblioteca de 2 kb, 100 kb ou 150 kb), a distância entre as leituras de extremidades era conhecida. Além disso, o alinhamento das sequências dos dois contigs por meio da utilização de leituras de pontas pareadas automaticamente determina a orientação relativa dos dois contigs. Desse modo, contigs de cópia única puderam ser unidos, embora com intervalos onde os elementos repetitivos estão localizados. Essas coleções espaçadas de sequências contigs unidas são denominadas **arcabouços** (por vezes também denominados **supercontigs**). Tendo em vista que a maior parte das repetições de *Drosophila* são grandes (3 a 8 kb) e amplamente espaçadas (uma repetição aproximadamente a cada 150 kb), essa técnica foi extremamente eficaz na produção de uma sequência rascunho corretamente montada do DNA de cópia única. Um resumo da lógica dessa abordagem está demonstrado na Figura 14.7.

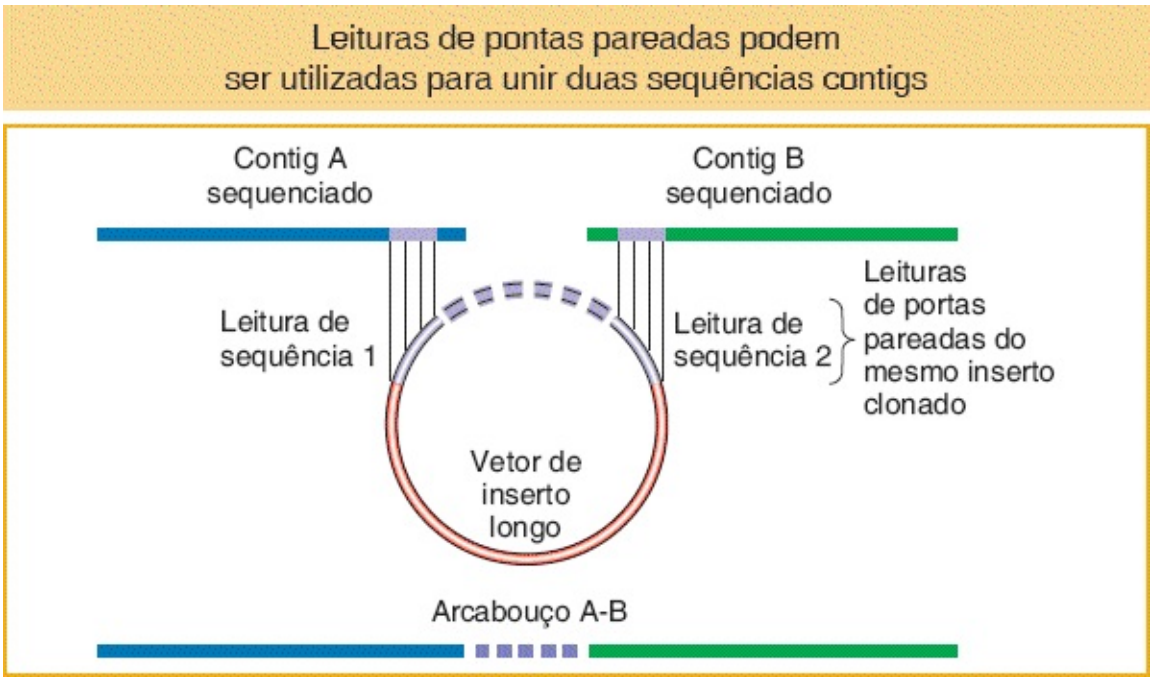


FIGURA 14.6 Leituras de pontas pareadas podem ser utilizadas para unir duas sequências contigs em um único arcabouço ordenado e orientado.

O WGS de nova geração não conseguiu contornar o problema das sequências repetitivas e dos intervalos. Tendo em vista que essa abordagem se destina a evitar a construção de bibliotecas, que por sua vez facilitaria a ligação dos intervalos entre os contigs por meio de leituras de pontas pareadas, pesquisadores do WGS de nova geração precisavam planejar um modo de ligar esses intervalos sem construir bibliotecas genômicas em vetores. Uma solução era construir uma biblioteca de fragmentos circulares de DNA genômico de tamanhos desejados. A circularização possibilita que segmentos curtos de sequências anteriormente distantes localizadas nas extremidades de cada fragmento sejam justapostas em cada lado de uma sequência ligadora. A obtenção dessas moléculas circulares, a amplificação e o sequenciamento dos fragmentos que contêm ligadores produzem leituras de pontas pareadas equivalentes àquelas obtidas a partir do sequenciamento de insertos de bibliotecas genômicas tradicionais (Figura 14.8).

Tanto no sequenciamento tradicional quanto no *shotgun* de genoma inteiro de última geração, normalmente permanecem alguns intervalos. Devem ser utilizados procedimentos específicos direcionados aos intervalos individuais para

preencher os dados ausentes nas montagens de sequências. Se os intervalos forem curtos, os fragmentos que faltam podem ser gerados por meio da utilização das sequências conhecidas nas extremidades das montagens como *primers* para amplificar e analisar a sequência genômica entre elas. Se os intervalos forem mais longos, podem ser realizadas tentativas para isolar as sequências ausentes como partes de insertos maiores que foram clonadas em um vetor e em seguida sequenciar os insertos.

Se um genoma é sequenciado nos padrões "rascunho" ou "finalizado" é uma questão de custo-benefício. É relativamente fácil criar um rascunho, mas muito difícil finalizar uma sequência.

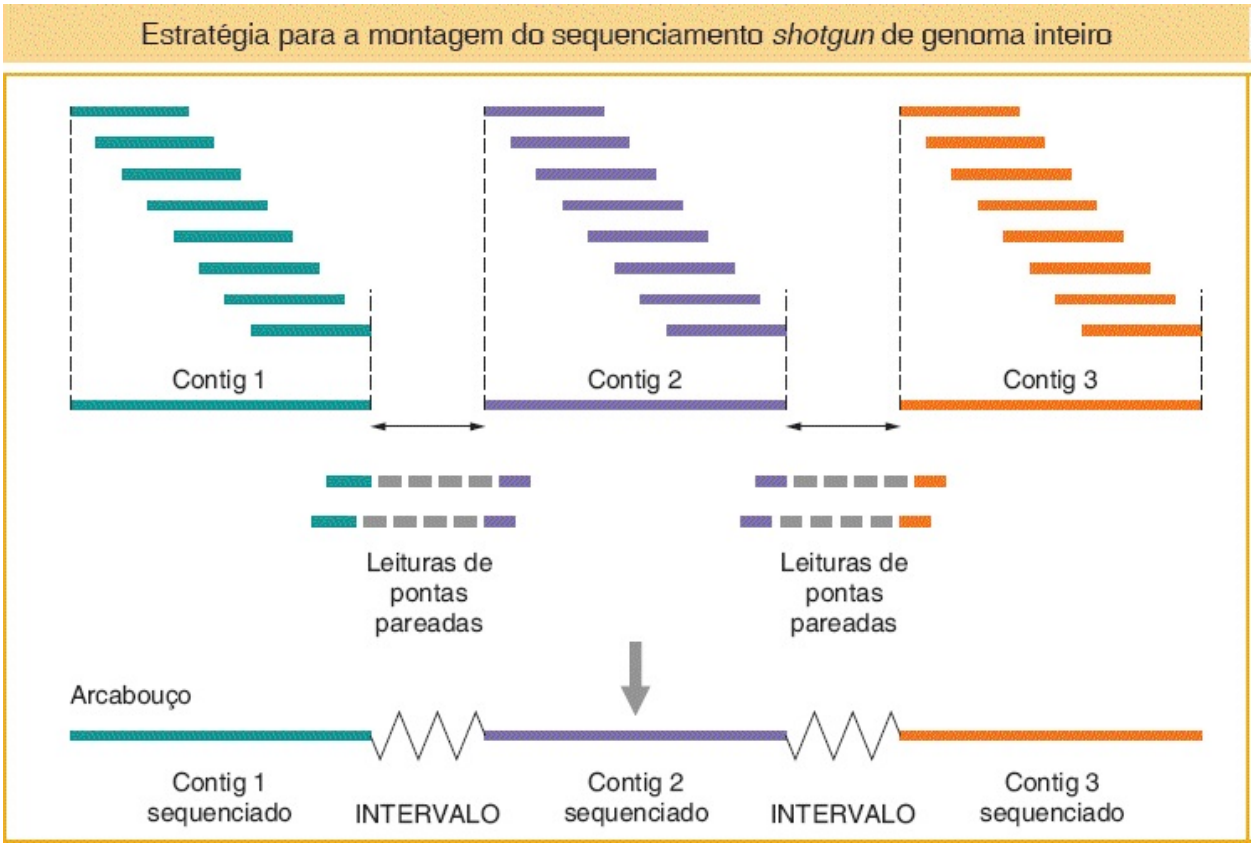


FIGURA 14.7 No sequenciamento *shotgun* de genoma inteiro, primeiramente as sobreposições de sequência única entre as leituras de sequência são utilizadas para construir contigs. Em seguida as leituras de pontas pareadas são utilizadas para abranger os intervalos e para ordenar e orientar os contigs em unidades maiores, denominadas *arcabouços*.

Leituras de pontas pareadas podem ser produzidas por meio de circularização

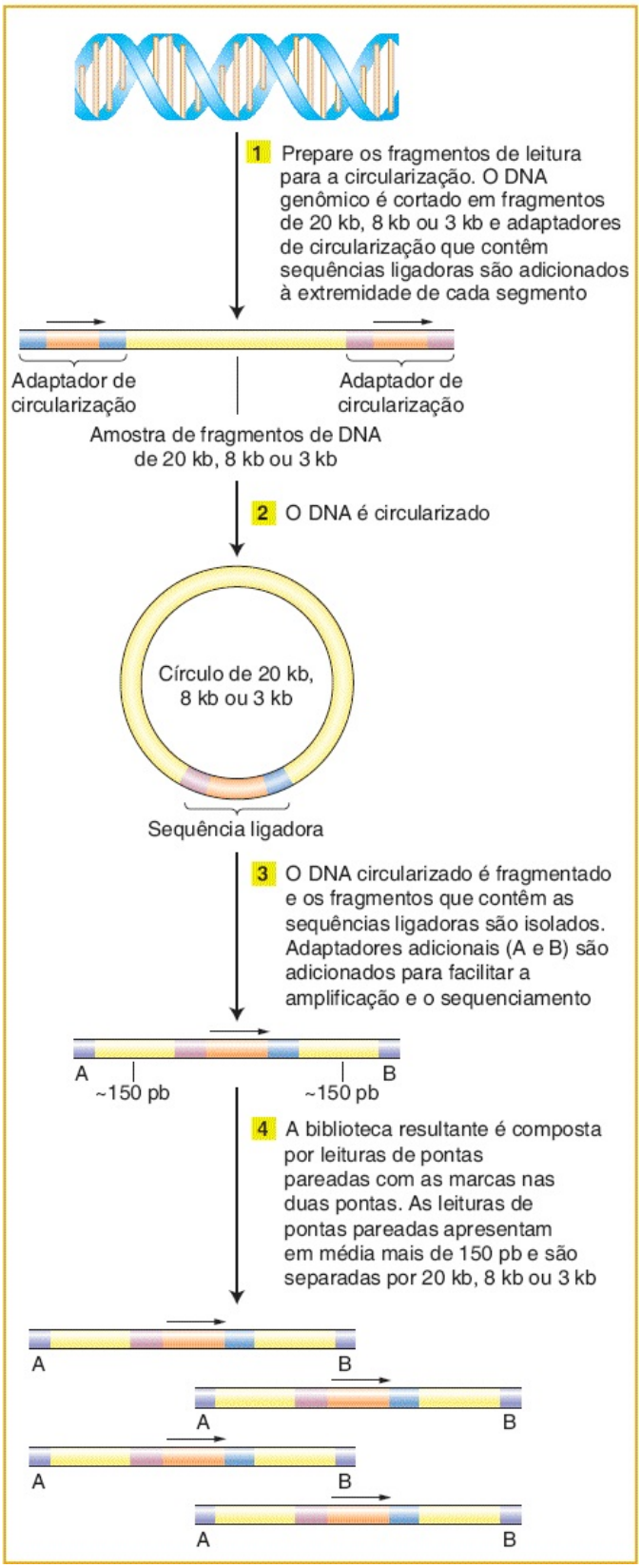


FIGURA 14.8 Leituras de pontas pareadas para o sequenciamento de alta produtividade podem ser produzidas sem a construção de bibliotecas genômicas. A figura tem por base o protocolo de ponta pareada da Roche GS FLX Titanium Series, Roche Applied Science, Mannheim, Alemanha.

14.3 Bioinformática | Significado da sequência genômica

A sequência genômica é um código altamente criptografado que contém a informação bruta para a construção e o funcionamento dos organismos. O estudo do conteúdo de informação dos genomas é denominado **bioinformática**. Estamos longe de ser capazes de ler essa informação do início ao fim, do modo como lemos um livro. Muito embora saibamos quais trincas codificam quais aminoácidos nos segmentos codificadores de proteínas, uma grande parte das informações contidas em um genoma não é decifrável a partir da mera inspeção.

Natureza do conteúdo da informação do DNA

O DNA contém informação, mas como ela é codificada? Convencionalmente, pensamos na informação como a soma de todos os produtos gênicos, proteínas e RNA. Entretanto, o conteúdo de informação do genoma é mais complexo do que isso. O genoma também contém sítios de ligação para diferentes proteínas e RNA. Muitas proteínas se ligam a sítios localizados no próprio DNA, enquanto outras proteínas e RNA se ligam a sítios localizados no mRNA (Figura 14.9). A sequência e as posições relativas dos sítios possibilitam que os genes sejam transcritos, recompostos e traduzidos adequadamente, na ocasião apropriada e no tecido apropriado. Por exemplo, sítios de ligação de proteínas reguladoras determinam quando, onde e em que nível um gene será expresso. No nível do RNA em eucariotos, as localizações dos sítios de ligação para os RNA e as proteínas dos spliceossomos determinarão os sítios de *splicing* em 5′ e 3′ nos quais os íntrons são removidos. Independentemente de um sítio de ligação de fato atuar como tal no DNA ou RNA, o sítio deve estar codificado no DNA. A informação no genoma pode ser considerada como a soma de todas as sequências que codificam proteínas e RNA, mais os sítios de ligação que regulam a ocasião e

o local de suas ações. Na medida em que o rascunho do genoma continua a ser melhorado, o objetivo principal é a identificação de todos os elementos funcionais do genoma. Esse processo é denominado **anotação**.

Dedução dos genes codificadores de proteínas a partir da sequência genômica

Tendo em vista que as proteínas presentes em uma célula determinam amplamente sua morfologia e propriedades fisiológicas, uma das primeiras prioridades na análise e na anotação do genoma é tentar determinar um inventário de todos os polipeptídios codificados pelo genoma de um organismo. Esse inventário é denominado **proteoma** do organismo. Ele pode ser considerado uma "lista parcial" para a célula. Para determinar a lista de polipeptídios, a sequência de cada mRNA codificado pelo genoma deve ser deduzida. Em virtude da remoção dos íntrons, essa tarefa é particularmente desafiadora em eucariotos multicelulares, nos quais os íntrons são a regra. Em seres humanos, por exemplo, um gene médio apresenta aproximadamente 10 éxons. Além disso, muitos genes codificam éxons alternativos; ou seja, alguns éxons são incluídos em algumas versões de um mRNA processado, mas não são incluídos em outras (ver Capítulo 8). Os mRNA alternativamente processados podem codificar polipeptídios que apresentam uma grande parte das suas sequências de aminoácidos em comum, porém não todas. Muito embora tenhamos muitos exemplos de genes e mRNA completamente sequenciados, ainda não conseguimos identificar os sítios de *splicing* 5′ e 3′ meramente a partir da sequência de DNA com um alto grau de precisão. Portanto, não podemos ter certeza sobre quais sequências são íntrons. As previsões de éxons utilizados alternativamente são ainda mais propensas a erros. Em virtude de tais motivos, a dedução da lista total de partes de polipeptídios em eucariotos superiores é um grande problema. Seguem algumas abordagens.

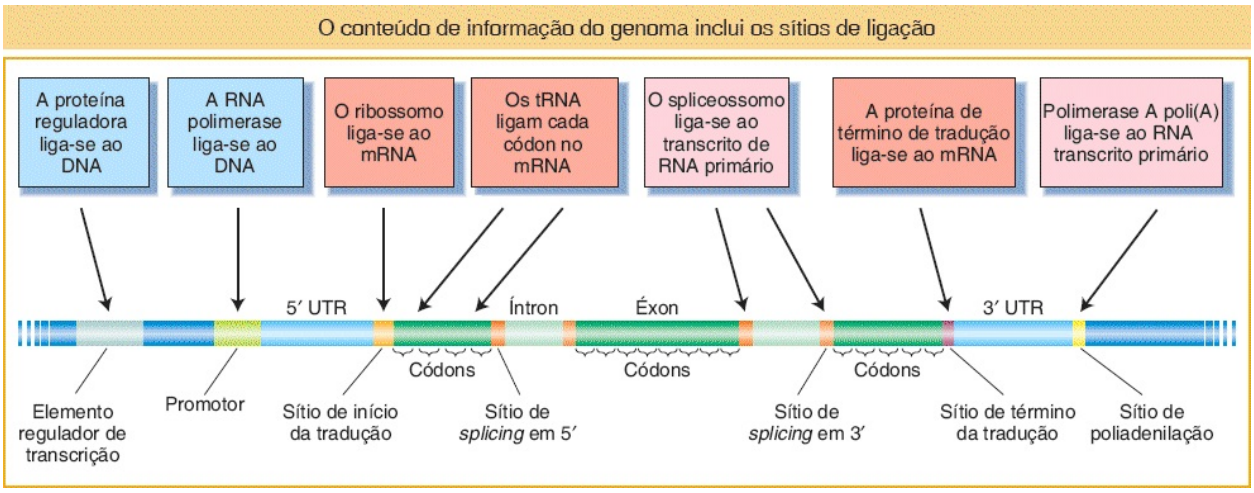


FIGURA 14.9 Um gene no DNA pode ser visualizado como uma série de sítios de ligação para as proteínas e RNA.

Detecção de ORF. A principal abordagem para a produção de uma lista de polipeptídios é utilizar a análise computadorizada da sequência do genoma para prever as sequências de mRNA e polipeptídios, uma parte importante da bioinformática. O procedimento é procurar por sequências que apresentem as características de genes. Essas sequências seriam do tamanho de um gene e compostas por códons com sentido após a remoção de possíveis íntrons. As sequências apropriadas das extremidade 5′ e 3′ estariam presentes, tais como os códons de início e de parada. As sequências com essas características típicas de genes são denominadas **matrizes de leitura abertas** (**ORF**, *open reading frames*). Para encontrar candidatos a ORF, programas de computador realizam a varredura da sequência do DNA em ambos os filamentos em cada matriz de leitura. Tendo em vista que existem três possíveis matrizes de leitura em cada filamento, existem seis matrizes de leitura possíveis no total.

Evidências diretas a partir de sequências de cDNA. Outro meio de identificar ORF e éxons é por meio da análise da expressão de mRNA. Essa análise pode ser realizada de dois modos. Ambos os métodos envolvem a síntese de bibliotecas de moléculas de DNA complementares a sequências de mRNA, denominadas cDNA (ver Capítulo 10). O método estabelecido mais longo envolve a clonagem e a amplificação dessas moléculas de cDNA em um vetor. Entretanto, tecnologias de

NGS possibilitam o sequenciamento direto de moléculas de cDNA curtas sem a etapa da clonagem (denominado **sequenciamento de RNA**, ou "RNA-seq", abreviadamente). Seja qual for o método utilizado, as sequências de DNA complementar são extremamente valiosas de dois modos. Primeiramente, elas são a evidência direta de que um determinado segmento do genoma é expresso e, portanto, pode codificar um gene. Em segundo lugar, em virtude de o cDNA ser complementar ao mRNA final, os íntrons do transcrito primário foram removidos, o que facilita muito a identificação dos éxons e íntrons de um gene (Figura 14.10).

O alinhamento dos cDNA com a sua sequência genômica correspondente delineia claramente os éxons e, portanto, os íntrons são revelados como as regiões que se situam entre os éxons. Na sequência de cDNA montada, o ORF deve ser contínuo desde o códon de início até o códon de fim. Portanto, as sequências de cDNA podem auxiliar muito na identificação da matriz de leitura correta, incluindo os códons de início e fim. Evidência de cDNA completo é tida como a comprovação padrão-ouro de que foi identificada a sequência de uma unidade de transcrição, incluindo seus éxons e a sua localização no genoma.

Além das sequências do cDNA completas, existem grandes conjuntos de dados de cDNA em relação aos quais apenas as extremidades 5′ ou 3′ ou ambas foram sequenciadas. Essas leituras de sequências de cDNA curtas são denominadas **etiquetas de sequências expressas (EST)**. As etiquetas de sequências expressas podem ser alinhadas com o DNA genômico e, assim, ser utilizadas para determinar as extremidades 5′ e 3′ dos transcritos — em outras palavras, para determinar os limites do transcrito, conforme demonstrado na Figura 14.10.

cDNA e EST revelam éxons ou extremidades de genes nas pesquisas genômicas

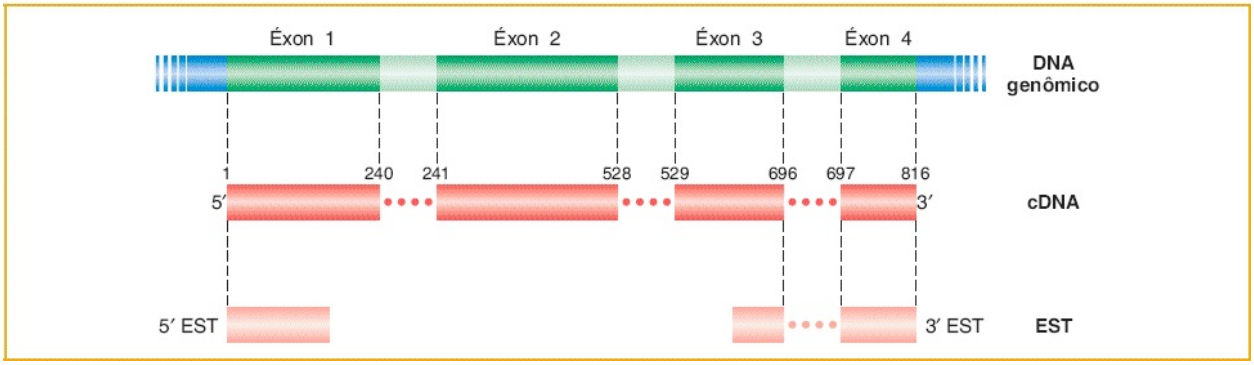


FIGURA 14.10 Alinhamento de DNA complementares (cDNA) totalmente sequenciados e de etiquetas de sequências expressas (EST) com DNA genômico. As *linhas tracejadas* indicam as regiões de alinhamento; em relação ao cDNA, estas regiões são os éxons do gene. Os *pontos* entre os segmentos de cDNA ou EST indicam regiões no DNA genômico que não estão alinhadas com as sequências de cDNA ou EST; essas regiões são locais de íntrons. Os números acima da linha do cDNA indicam as coordenadas das bases da sequência de cDNA, em que a base 1 é a base mais próxima de 5′ e a base 816 é a mais próxima de 3′ do cDNA. Em relação às EST, apenas uma leitura de sequência curta é obtida a partir de cada extremidade (5′ e 3′) do cDNA correspondente. Essas leituras de sequências estabelecem os limites da unidade de transcrição, mas não são informativas a respeito da estrutura interna do transcrito, exceto se as sequências EST cruzarem um íntron (como é verdadeiro para a 3′ EST ilustrada aqui).

Previsões dos sítios de ligação. Conforme já discutido, um gene consiste em um segmento de DNA que codifica um transcrito, bem como em sinais reguladores que determinam quando, onde e em que quantidade aquele transcrito é produzido. Por sua vez, esse transcrito apresenta os sinais necessários para determinar a sua recomposição em mRNA e a tradução desse mRNA em um polipeptídio (Figura 14.11). Atualmente, existem programas de computador estatísticos de "busca de genes", os quais procuram por sequências previstas dos diversos sítios de ligação utilizados como promotores, por sítios de início de transcrição, sítios de *splicing* 3′ e 5′ e por códons de início de tradução no DNA genômico. Essas previsões baseiam-se em *motifs* consenso para as referidas sequências conhecidas, mas não são perfeitas.

Uso de polipeptídios e similaridade de DNA. Tendo em vista que os organismos apresentam ancestrais em comum, eles também apresentam muitos genes com sequências similares em comum. Portanto, um gene provavelmente apresentará correlatos entre os genes isolados e sequenciados em outros organismos, especialmente nos muito correlatos. Os genes candidatos previstos pelas técnicas precedentes com frequência podem ser verificados por meio de sua comparação com todas as outras sequências de genes que já foram encontradas. Uma sequência candidata é apresentada como uma "sequência indagada" às bases de dados públicas que contêm um registro de todas as sequências gênicas conhecidas. Esse procedimento é denominado pesquisa BLAST (Basic Local Alignment Search Tool). A sequência pode ser apresentada como uma sequência de nucleotídios (uma pesquisa BLASTn) ou como uma sequência traduzida de

aminoácidos (BLASTp). O computador realiza a varredura da base de dados e fornece uma lista parcial ou total de "alvos", iniciando com as correspondências mais próximas. Se a sequência candidata se assemelhar de modo muito próximo àquela de um gene anteriormente identificado de outro organismo, então essa semelhança proporciona uma forte indicação de que o gene candidato é o gene de fato. Correspondências menos próximas também são úteis. Por exemplo, uma identidade de aminoácidos de apenas 35%, mas em posições idênticas, é um forte indicador de que duas proteínas apresentam uma estrutura tridimensional comum.

As pesquisas BLAST são utilizadas de muitos outros modos, mas o objetivo sempre é saber mais a respeito de alguma sequência de interesse identificada.

Previsões com base na tendenciosidade de códons. Lembre, do Capítulo 9, que o código triplo para aminoácidos é degenerado; ou seja, a maior parte dos aminoácidos é codificada por dois ou mais códons (ver Figura 9.5). Os códons múltiplos para um único aminoácido são denominados *códons sinônimos*. Em uma determinada espécie, nem todos os códons sinônimos para um aminoácido são utilizados com frequência igual. Em vez disso, determinados códons estão presentes muito mais frequentemente nos mRNA (e, portanto, no DNA que os codifica). Por exemplo, em *D. melanogaster*, dos dois códons para cisteína, UGC é utilizado em 73% das ocasiões, enquanto UGU é utilizado em 27% das ocasiões. Essa utilização é um diagnóstico de *Drosophila*, uma vez que, em outros organismos, esse padrão de "tendenciosidade de códons" é razoavelmente diferente. Acredita-se que as tendenciosidades de códons ocorram em virtude da abundância relativa dos tRNA complementares a esses diversos códons em uma determinada espécie. Se a utilização do códon de um ORF previsto corresponde ao padrão conhecido de utilização de códons daquela espécie, então essa correspondência é uma evidência que ampara que o ORF proposto é genuíno.

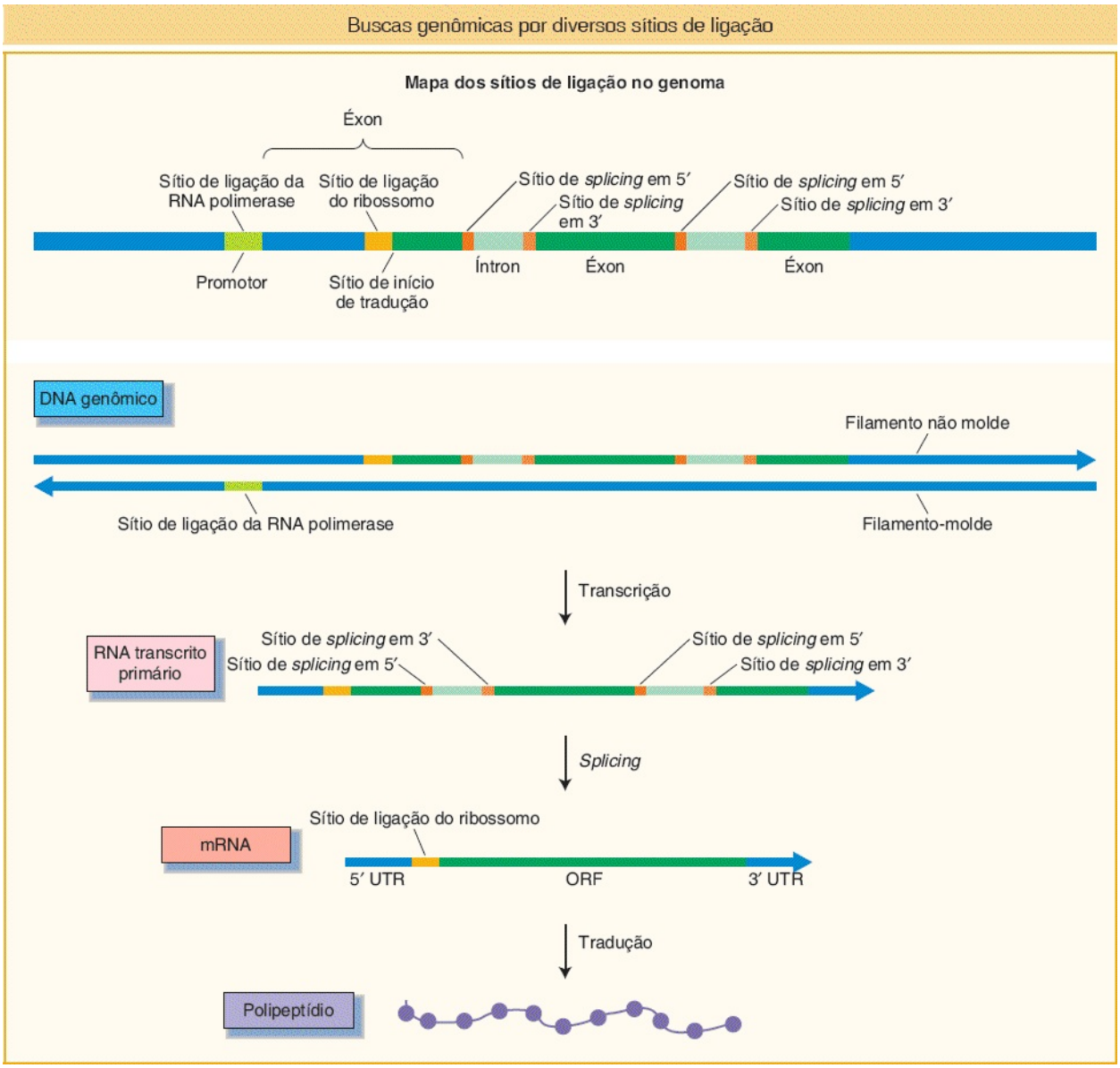


FIGURA 14.11 Transferência de informação eucariótica de um gene para uma cadeia polipeptídica. Observe os "sítios de ligação" do DNA e RNA que estão conectados por complexos proteicos para iniciar os eventos de transcrição, *splicing* e tradução.

Resumo. Um resumo de como diferentes fontes de informação são combinadas para criar as melhores previsões possíveis do mRNA e dos genes está ilustrado na Figura 14.12. Esses diferentes tipos de evidências são complementares e podem ter validação cruzada entre si. Por exemplo, a estrutura de um gene pode ser inferida a partir da evidência da similaridade da proteína dentro de uma região do DNA genômico ligado pelas EST 5′ e 3′. São possíveis previsões úteis

até mesmo sem uma sequência de cDNA ou evidência de similaridade proteica. Um programa de previsão de sítio de ligação pode propor um ORF hipotético e a tendenciosidade de códon adequado seria a evidência de suporte.

> CONCEITO-CHAVE As previsões da estrutura polipeptídica e do mRNA a partir da sequência do DNA genômico dependem da integração da informação da sequência de cDNA, das previsões do sítio de ligação, das similaridades de polipeptídios e da tendenciosidade de códons.

Consideraremos algumas das percepções de nossa primeira visão das estruturas genômicas em geral e de listas de partes globais de algumas espécies cujos genomas foram sequenciados. Iniciaremos conosco mesmos. O que podemos aprender ao observar o genoma humano por si próprio? Em seguida, veremos o que podemos aprender ao comparar nosso genoma com outros.

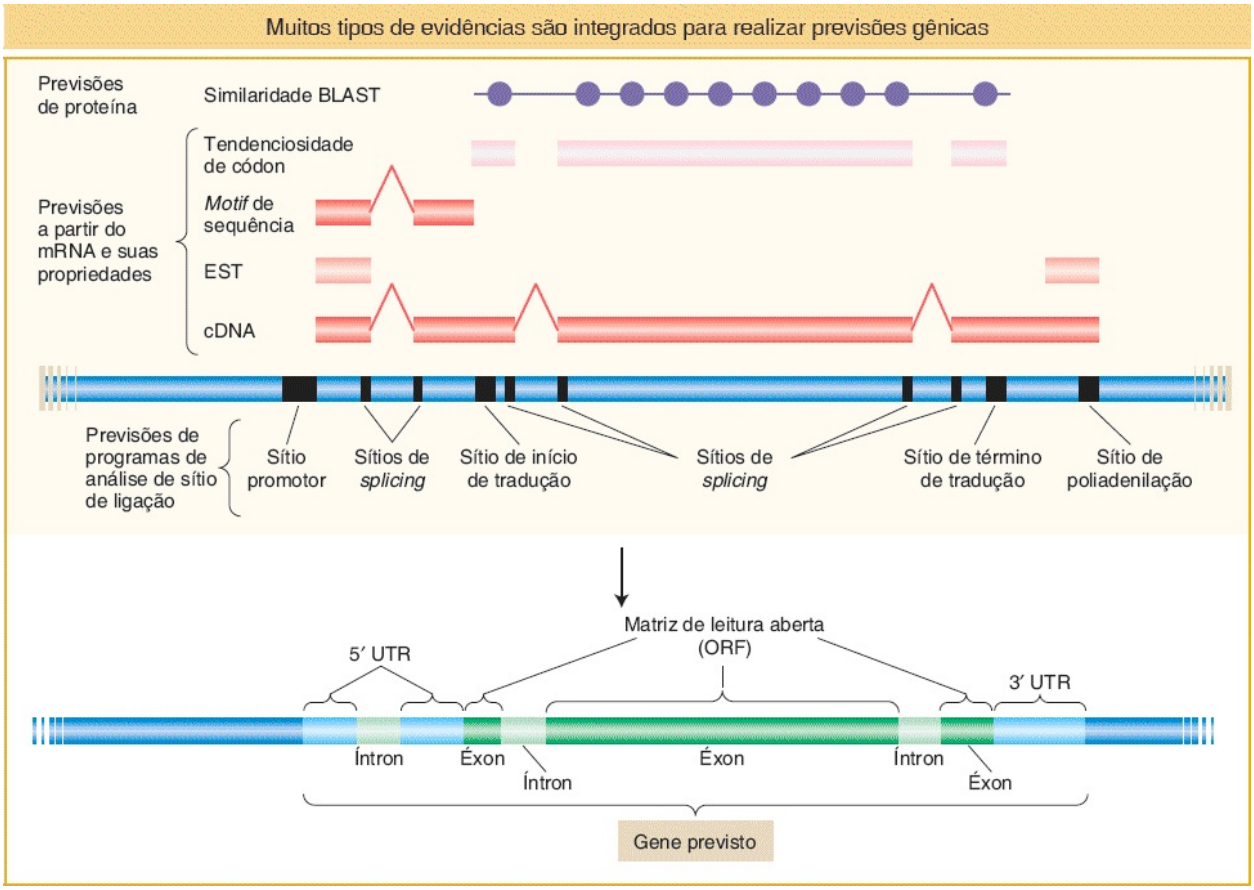


FIGURA 14.12 Os diferentes modos de evidenciar produtos gênicos — cDNA, EST, correspondências similares de BLAST, tendenciosidade de códons e concordância de *motifs* — são integrados para realizar previsões sobre os genes. Onde se observa que múltiplas classes de evidências estão associadas a uma sequência de DNA genômico em particular, há maior confiança na probabilidade de que uma previsão gênica seja precisa.

14.4 Estrutura do genoma humano

Ao descrever a estrutura geral do genoma humano, primeiramente devemos confrontar sua estrutura repetida. Uma fração considerável do genoma humano, aproximadamente 45%, é repetitiva. Uma grande parte desse DNA repetitivo é composta por cópias de elementos de transposição. De fato, até mesmo no DNA de cópia única restante, uma fração apresenta sequências que sugerem que elas podem ser descendentes de elementos de transposição antigos que agora são imóveis e acumularam mutações aleatórias, causando a sua divergência na sequência em relação aos elementos de transposição ancestrais. Portanto, uma grande parte do genoma humano parece ser composta por "caronas" genéticos.

Apenas uma pequena parte do genoma humano codifica polipeptídios; ou seja, um pouco menos de 3% dele codifica éxons dos mRNA. Os éxons são tipicamente pequenos (aproximadamente 150 bases), enquanto os íntrons são grandes, muitos se estendendo a mais de 1.000 bases e alguns se estendendo a mais de 100.000 bases. Os transcritos são compostos por uma média de 10 éxons, embora muitos apresentem substancialmente mais. Finalmente, os íntrons podem ser submetidos a *splicing* do mesmo gene em locais variáveis. Essa variação na localização dos sítios de *splicing* gera diversidade adicional considerável ao mRNA e à sequência polipeptídica. Com base nos atuais dados de cDNA e EST, no mínimo 60% dos genes humanos codificadores de proteínas provavelmente apresentam duas ou mais variantes de corte. Em média, existem diversas variantes de corte por gene. Portanto, o número de proteínas distintas codificadas pelo genoma humano é muito superior ao número de genes reconhecidos.

Não foi fácil definir o número de genes no genoma humano. No rascunho inicial do genoma humano, havia uma estimativa de 30.000 a 40.000 genes codificadores de proteínas. Entretanto, a complexa arquitetura desses genes e do genoma pode

tornar a anotação difícil. Algumas sequências registradas como genes realmente podem ser éxons de genes maiores. Além disso, existem mais de 19.000 **pseudogenes**, que são ORF ou ORF parciais que primeiramente podem parecer genes, mas não são funcionais ou são inativos em virtude do modo de sua origem ou de mutações. Os assim denominados **pseudogenes processados** são sequências de DNA que foram transcritas de modo reverso a partir do RNA e inseridas aleatoriamente no genoma. Aproximadamente 90% dos pseudogenes humanos aparentam ser desse tipo. Aproximadamente 900 pseudogenes parecem genes convencionais que adquiriram uma ou mais mutações que romperam ORF no decorrer da evolução. Na medida em que os desafios na anotação foram superados, o número estimado de genes no genoma humano diminuiu drasticamente. Uma estimativa recente é que existam aproximadamente 21.000 genes codificadores de proteínas.

A anotação do genoma humano progrediu na medida em que as sequências de cada cromossomo foram finalizadas uma a uma. Essas sequências em seguida se tornaram a área de pesquisas na busca por genes candidatos. Em exemplo de previsões de genes para um cromossomo do genoma humano está demonstrado na Figura 14.13. As referidas previsões estão sendo continuamente revisadas na medida em que novos dados se tornam disponíveis. O atual estado das previsões pode ser visualizado em diversos *sites*, mais notavelmente nas bases de dados de DNA públicas nos EUA e na Europa (ver Apêndice B). Essas previsões são as melhores inferências atuais sobre os genes codificadores de proteínas presentes nas espécies sequenciadas e, como tal, são trabalhos em andamento.

Elementos funcionais não codificadores no genoma

A discussão até agora focou exclusivamente nas regiões do genoma que codificam proteínas. Essa ênfase ocorre mais em virtude da facilidade analítica do que da importância biológica. Em virtude da simplicidade e da universalidade do código genético e da capacidade de sintetizar cDNA a partir de mRNA, a detecção de ORF e éxons é muito mais fácil do que a detecção de sequências não codificadoras funcionais. Conforme declarado anteriormente, apenas 3% do genoma humano codificam éxons de mRNA e menos da metade dessas sequências

de éxons, pouco mais de 1% do DNA total do genoma, codifica sequências proteicas. Assim, mais de 98% do nosso genoma não codifica proteínas. Como identificamos outras partes funcionais do genoma?

Os íntrons e as sequências 5′ e 3′ não traduzidas são prontamente anotados por meio da análise de transcritos gênicos, enquanto os promotores dos genes normalmente são identificados por meio de sua proximidade das unidades de transcrição e de sequências sinalizadoras no DNA. Entretanto, outras sequências reguladoras, tais como acentuadores, não são identificáveis por meio da mera inspeção de sequências de DNA e outras sequências que codificam diversos tipos de transcritos de RNA (microRNA, pequenos RNA de interferência, longos RNA não codificadores) requerem a detecção e a anotação de seus transcritos. Embora muitos dos referidos elementos não codificadores tenham sido identificados no curso do estudo da genética molecular humana, o potencialmente vasto número dos elementos recomenda uma abordagem mais sistemática. O projeto Encyclopedia of DNA Elements (ENCODE), portanto, foi lançado com o ambicioso objetivo de identificar todos os elementos funcionais do genoma humano.

Esse esforço cooperativo em grande escala empregou um arranjo diverso de técnicas para detectar sequências possivelmente envolvidas no controle da transcrição gênica, bem como de todas as regiões transcritas. Tendo em vista que se espera que tais sequências sejam ativas apenas em tipos celulares individuais ou subconjuntos de tipos celulares, pesquisadores estudaram 147 tipos celulares humanos. Ao procurar regiões que estivessem associadas à ligação de fatores de transcrição, o projeto ENCODE estimou que existam aproximadamente 500.000 possíveis acentuadores associados aos genes conhecidos. O projeto também detectou transcritos que emanam de 80% do genoma humano.

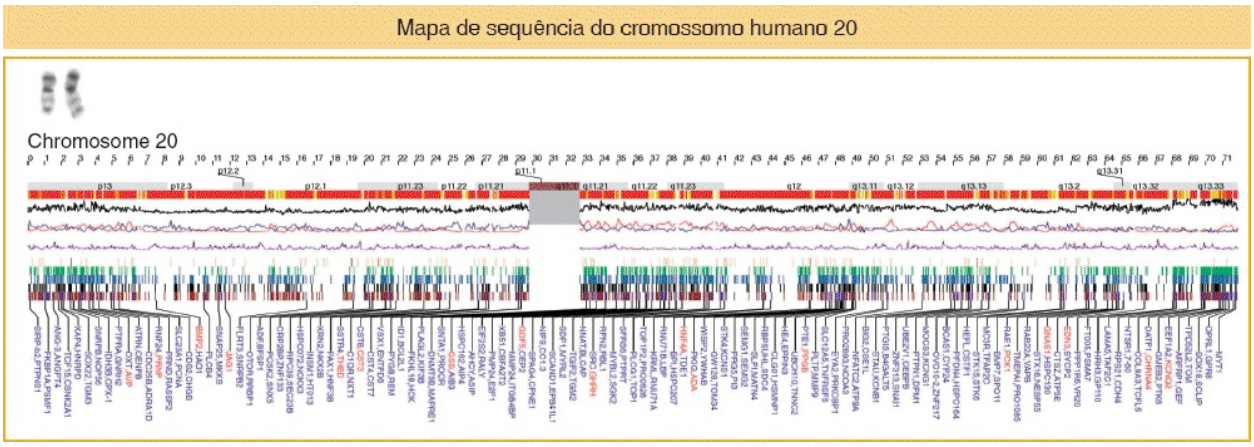


FIGURA 14.13 Foram identificados diversos genes no cromossomo humano 20. As coordenadas do mapa de recombinação e citogenético estão demonstradas nas linhas superiores da figura. Diversos gráficos que ilustram a densidade gênica e as diferentes propriedades do DNA estão demonstrados nas seções intermediárias. Os identificadores dos genes previstos estão demonstrados na parte inferior do painel. (Cortesia de Jim Kent, Ewan Birney, Darryl Leja e Francis Collins. De acordo com o International Human Genome Sequencing Consortium, "Initial Sequencing and Analysis of the Human Genome", *Nature* 409, 2001, 860-921.)

Essa é uma fração muito maior do genoma do que era esperado. Afinal, conforme declarado anteriormente, apenas um pouco mais de 1% do genoma é de sequências codificadoras. Entretanto, a produção de um transcrito não necessariamente significa que o transcrito contribui para a biologia humana. É possível que alguma proporção desses transcritos represente "ruídos" na célula — transcritos que não apresentam função biológica, mas que também não prejudicam. Não é sensato atribuir uma função a uma sequência sem algum tipo de dados adicionais. Assim, quais tipos de dados adicionais podem ser utilizados para resolver questões de função?

A conservação evolutiva de sequências comprovou ser um bom indicador da função biológica. Sequências não serão conservadas ao longo do período evolutivo, exceto se mutações que as alteram forem eliminadas por seleção natural. Um modo de localizar elementos não codificadores potencialmente funcionais, então, é procurar por sequências conservadas, que não foram alteradas ao longo de milhões de anos de evolução.

Por exemplo, pode-se procurar sequências altamente conservadas de comprimento modesto entre algumas espécies ou sequências menos perfeitamente

conservadas de maior comprimento entre um número maior de espécies. Comparações dos genomas de seres humanos, ratos e camundongos levaram à identificação dos assim denominados *elementos ultraconservados*, sequências perfeitamente conservadas entre as três espécies. Pesquisas nesses genomas identificaram mais de 5.000 sequências de mais de 100 pb e 481 sequências de mais de 200 pb que são absolutamente conservadas. Embora muitos desses elementos sejam observados em regiões pobres em genes, eles são mais ricamente concentrados próximos a genes reguladores importantes para o desenvolvimento. A maioria dos elementos não codificadores altamente conservados pode participar amplamente na regulação da expressão da *toolkit* genética para o desenvolvimento de mamíferos e outros vertebrados (ver Capítulo 13).

Como podemos verificar que os referidos elementos conservados desempenham um papel na regulação gênica? Esses elementos podem ser testados do mesmo modo que os elementos reguladores de transcrição de ação cis examinados nos capítulos anteriores, com a utilização de genes repórter (ver Seção 12.2). Um pesquisador posiciona as regiões reguladoras candidatas adjacentes a um promotor e ao gene repórter e introduz o gene repórter em uma espécie hospedeira. Um referido exemplo está demonstrado na Figura 14.14. Um elemento que é altamente conservado entre mamíferos, galinhas e uma espécie de rã está localizado a 488 kb da extremidade 3′ do gene *ISL1* humano, que codifica uma proteína necessária para a diferenciação de neurônio motor. Esse elemento foi posicionado *upstream* de um promotor e do gene repórter de β-galactosidade (*lacZ*) e o construto foi injetado em pronúcleos de ovócitos fertilizados de camundongos. Em seguida observa-se se proteína repórter é expressa ao longo da medula espinal e na cabeça, como seria esperado em relação à localização de futuros neurônios motores (ver Figura 14.14). Mais significativamente, o padrão de expressão corresponde a parte do padrão de expressão do gene *ISL1* nativo de camundongo (presumidamente outros elementos não codificadores controlam as outras características da expressão de *ISL1*). O padrão de expressão sugere fortemente que o elemento conservado seja uma região reguladora do gene *ISL1* em cada espécie. Muitos milhares de elementos reguladores não codificadores

humanos provavelmente serão identificados com base na conservação de sequência e na atividade desses elementos em ensaios de repórter.

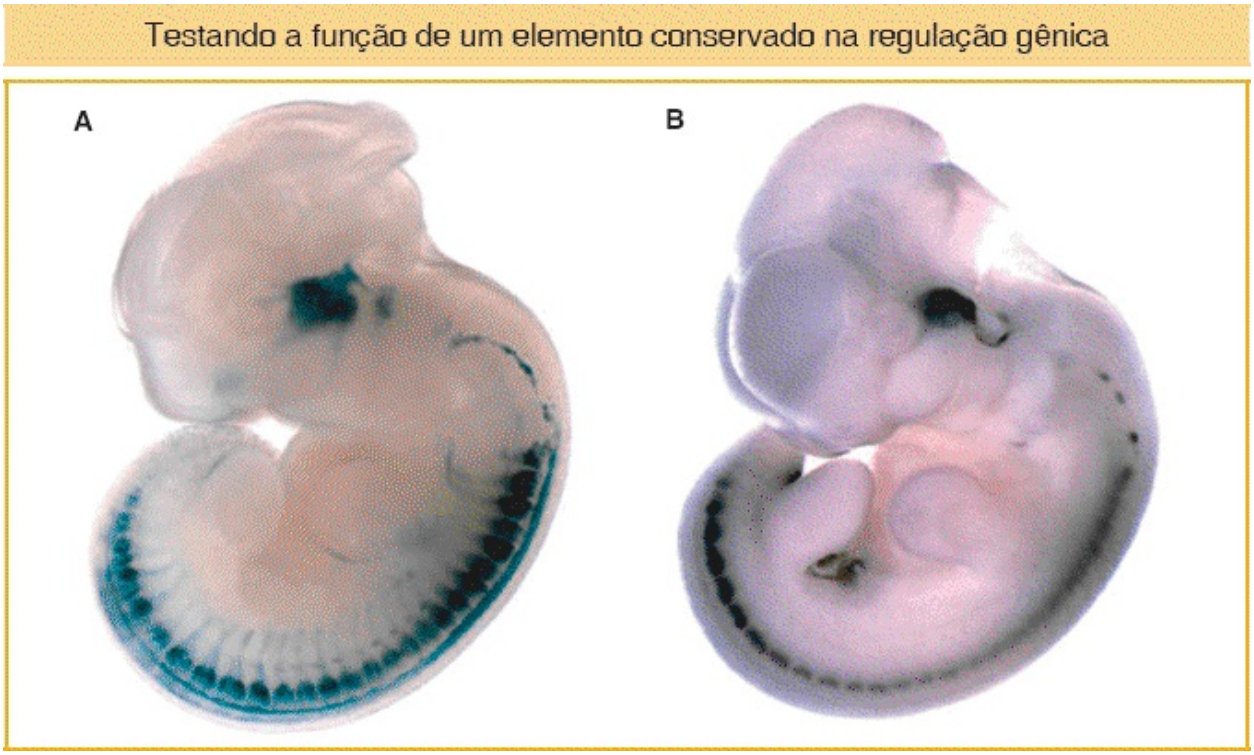


FIGURA 14.14 Um elemento regulador transcricional de ação cis é identificado em um elemento ultraconservado do genoma humano. Um elemento ultraconservado que está localizado próximo do gene *ISL1* humano foi acoplado a um gene repórter e injetado em ovócitos fertilizados de camundongo. As regiões nas quais o gene é expresso estão coradas em *azul-escuro* ou *preto*. **A.** O gene repórter é expresso na cabeça e na medula espinal de um camundongo transgênico, conforme observado aqui no dia 11,5 de gestação. Este padrão de expressão corresponde (**B**) ao padrão nativo de expressão do gene *ISL1* de camundongos no dia 11,5 de gestação. Este experimento demonstra como elementos não codificadores funcionais podem ser identificados por meio de genômica comparativa e testados em um organismo-modelo. (De G. Bejerono *et al.*, "A Distal Enhancer and an Ultraconserved Exon Are Derived from a Novel Retroposon", *Nature* 441, 2006, 87-90, Fig. 3.)

CONCEITO-CHAVE Os elementos funcionais não codificadores do genoma são muito mais difíceis de identificar do que as sequências codificadoras e requerem uma combinação de evidências comparativas e experimentais para a validação.

14.5 Genômica comparativa de seres humanos com outras espécies

Fundamentalmente, uma grande parte da ciência genômica envolve uma abordagem comparativa. Por exemplo, a maior parte do que sabemos a respeito da função das proteínas humanas é baseada na função de tais proteínas conforme analisadas em espécies-modelo. E muitas das questões que podem ser abordadas por meio da genômica são comparativas. Por exemplo, com frequência desejamos saber, assim como no caso de Nicholas Volker, como um indivíduo com um traço ou uma doença difere geneticamente daqueles sem eles.

A **genômica comparativa** também apresenta o potencial de revelar como as espécies divergem. As espécies evoluem e os traços são modificados por meio de alterações na sequência de DNA. O genoma, portanto, contém um registro da história evolutiva de uma espécie. Comparações entre os genomas das espécies podem revelar eventos únicos de linhagens em particular que podem contribuir para as diferenças na fisiologia, no comportamento ou na anatomia. Os referidos eventos podem incluir, por exemplo, o ganho e a perda de genes individuais ou de grupos de genes. Aqui, exploraremos os princípios-chave subjacentes à genômica comparativa e veremos alguns exemplos de como as comparações revelam o que é semelhante e diferente entre os seres humanos e outras espécies. Na próxima seção examinaremos como as diferenças são identificadas entre seres humanos individuais.

Inferência filogenética

A primeira etapa na comparação de genomas das espécies é decidir quais espécies devem ser comparadas. Para que as comparações sejam informativas, é crucial compreender as relações evolutivas entre as espécies a serem comparadas. A história evolutiva de um grupo é denominada **filogenia**. As filogenias são úteis por nos possibilitarem inferir como os genomas das espécies foram alterados ao longo do tempo.

A segunda etapa na comparação de genomas é a identificação dos genes

relacionados de modo mais próximo, denominados **homólogos**. Os genes homólogos podem ser reconhecidos por meio de semelhanças em suas sequências de DNA e de aminoácidos das proteínas que codificam. É importante distinguir aqui duas classes de genes homólogos. Alguns homólogos são genes no mesmo *locus* genético em diferentes espécies. Esses genes teriam sido herdados de um ancestral comum e são denominados **ortólogos**. Entretanto, muitos genes homólogos pertencem a famílias que foram expandidas (e contraídas) em número no período da evolução. Esses genes homólogos estão em diferentes *loci* genéticos no mesmo organismo. Eles tiveram origem quando genes em um genoma foram duplicados. Os genes relacionados por meio de eventos de duplicação gênica em um genoma são denominados **parálogos**. A história das famílias de genes pode ser razoavelmente reveladora a respeito da história evolutiva de um grupo.

Por exemplo, suponha que desejamos saber como o genoma de mamíferos evoluiu ao longo da história do grupo. Desejamos saber se os mamíferos como um grupo podem ter adquirido alguns genes únicos, se mamíferos com diferentes estilos de vida podem possuir conjuntos diferentes de genes e qual foi o destino dos genes que existiam em ancestrais mamíferos.

Felizmente, possuímos atualmente um conjunto grande e em expansão de sequências de genomas de mamíferos para comparar, que incluem representantes dos três principais ramos dos mamíferos — monotremados (p. ex., ornitorrinco), marsupiais (p. ex., canguru, gambá) e eutérios (p. ex., ser humano, cão, gato, camundongo). As relações entre esses grupos, alguns membros dentro desses grupos e outros vertebrados amniotas (amniotas são principalmente vertebrados terrestres que apresentam um ovo adaptado à vida terrestre) estão demonstradas na Figura 14.15.

Para ilustrar a importância da compreensão das filogenias e como as utilizar, consideramos o genoma do ornitorrinco. Os monotremados diferem de outros mamíferos pois põem ovos. A inspeção do genoma do ornitorrinco revelou que ele contém um gene para gema de ovo denominado vitelogenina. Análises dos genomas de marsupiais e eutérios não revelaram tal gene funcional. A presença de vitelogenina no ornitorrinco e sua ausência em outros mamíferos pode ser

explicada de dois modos: (1) a vitelogenina é uma invenção nova do ornitorrinco ou (2) a vitelogenina existia em um ancestral comum de monotremados, marsupiais e eutérios, mas foi subsequentemente perdida em marsupiais e eutérios. O sentido da alteração evolutiva é oposto nessas duas alternativas.

Uma comparação pareada simples entre o ornitorrinco e outro mamífero não distingue entre essas alternativas. Para tanto, primeiramente devemos inferir se a vitelogenina provavelmente estava presente no último ancestral comum do ornitorrinco, dos marsupiais e eutérios. Realizamos essa **inferência filogenética** ao examinar se a vitelogenina é observada fora desse grupo inteiro de mamíferos, que é denominado um **grupo evolutivo externo**. De fato, existem três genes de vitelogenina homólogos em galinha. Em seguida, consideramos a relação da galinha com os mamíferos. As galinhas pertencem a outro ramo importante dos amniotas. Observando a árvore evolutiva na Figura 14.15, podemos explicar a presença de vitelogeninas em galinhas e no ornitorrinco como o resultado de duas aquisições independentes (na linhagem do ornitorrinco e na linhagem da galinha, respectivamente) ou de apenas uma aquisição em um ancestral comum do ornitorrinco e da galinha (que, com base na árvore, seria um ancestral comum de todos os amniotas), seguida pela perda dos genes de vitelogenina em marsupiais e eutérios.

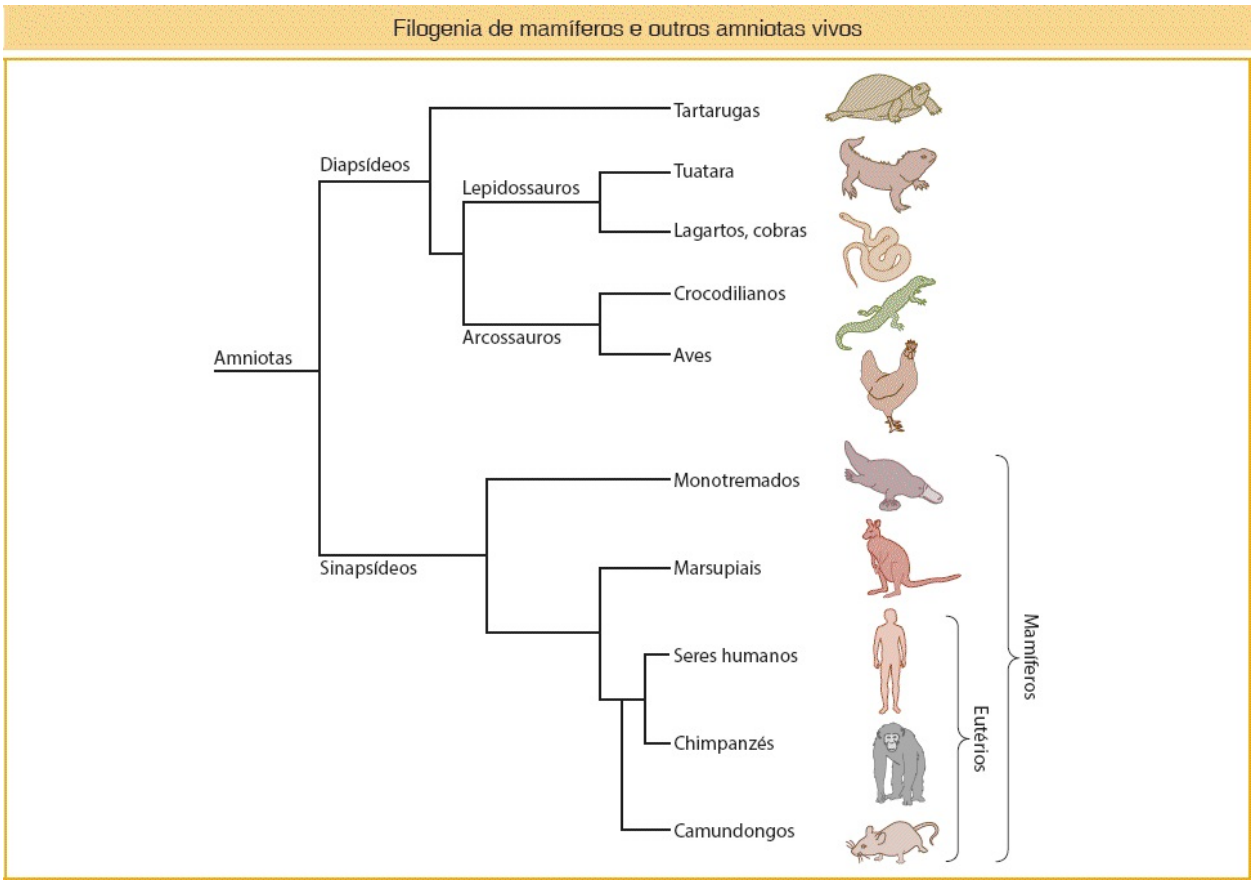


FIGURA 14.15 Filogenia de mamíferos e outros amniotas vivos. A árvore filogenética ilustra as relações evolutivas entre os três principais grupos de mamíferos (monotremados, marsupiais e eutérios) e outros amniotas, incluindo aves e diversos répteis. Por meio do mapeamento da presença ou da ausência de genes em grupos particulares de filogenia conhecida, pode-se inferir o sentido da alteração evolutiva (ganho ou perda) em linhagens particulares.

Como decidimos entre essas alternativas? Ao estudar eventos infrequentes, tais como a invenção de um gene, biólogos evolutivos preferem confiar no princípio da **parcimônia**, ou seja, favorecer a explicação mais simples que envolve o menor número de alterações evolutivas. Portanto, a explicação preferida para o padrão da evolução da vitelogenina em mamíferos é que essa proteína da gema do ovo e o gene correspondente estavam presentes em algum ancestral amniota ovíparo e foram retidos no ornitorrinco ovíparo e perdidos em animais não ovíparos.

Como é o caso, existe uma evidência adicional e muito persuasiva que ampara essa inferência. Embora a inspeção dos genomas de eutérios não revele quaisquer

genes da vitelogenina funcionais intactos, existem traços de sequências do gene da vitelogenina nos genomas de seres humanos e cães na mesma posição (sintênicos) dos genes da vitelogenina do ornitorrinco e da galinha (Figura 14.16). Essas sequências são relíquias moleculares de nossos ancestrais ovíparos. Na medida em que nossos ancestrais mamíferos se afastaram dos ovos com gemas, a seleção natural foi relaxada nas sequências do gene da vitelogenina, de tal modo que elas foram quase eliminadas por mutações ao longo de dezenas de milhões de anos. Nosso genoma contém diversas relíquias de genes que chegaram a ser funcionais em nossos ancestrais e, conforme veremos novamente nesta seção, as identidades desses pseudogenes refletem como a biologia humana divergiu da de nossos ancestrais.

É claro que a evolução também diz respeito à aquisição de novos traços. Por exemplo, a produção de leite é um traço compartilhado entre todos os mamíferos. Uma família de genes que codificam as proteínas caseína do leite é única dos mamíferos e está agrupada firmemente nos seus genomas, incluindo o do ornitorrinco. Apenas esse breve vislumbre de alguns poucos genomas de mamíferos nos informa que, sim, de fato, alguns mamíferos apresentam genes que outros não apresentam, alguns genes são compartilhados por todos os mamíferos e a presença ou a ausência de determinados genes está correlacionada com o estilo de vida dos mamíferos. O último é um achado persuasivo na genômica comparativa.

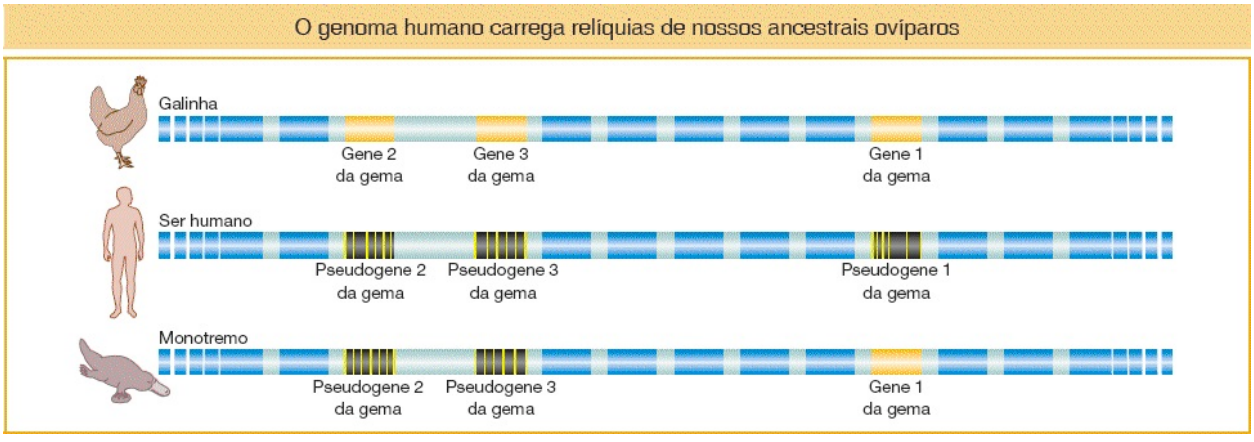


FIGURA 14.16 Sequências de genes ao longo do cromossomo 8 de galinha e do cromossomo 1 do ser humano e do ornitorrinco estão na mesma ordem relativa (*quadros*). Embora o genoma da galinha apresente

três genes que codificam as proteínas da gema do ovo, o ornitorrinco, ovíparo, apresenta um gene funcional e dois pseudogenes, e os seres humanos apresentam remanescentes muito curtos e fragmentados dos genes da gema.

> CONCEITO-CHAVE A determinação de quais elementos genômicos foram adquiridos ou perdidos durante a evolução requer o conhecimento sobre a filogenia das espécies que estão sendo comparadas. A presença ou a ausência de genes com frequência está correlacionada com os estilos de vida do organismo.

Vejamos mais alguns exemplos que iluminam a história evolutiva de nosso genoma e como somos diferentes de e semelhantes a outros mamíferos.

Camundongos e humanos

A sequência do genoma de camundongo tem sido particularmente informativa para a compreensão do genoma humano em virtude do papel de longa data do camundongo como uma espécie-modelo genética, do vasto conhecimento sobre a sua genética clássica e da relação evolutiva do camundongo com os seres humanos. As linhagens humanas e de camundongos divergiram há aproximadamente 75 milhões de anos, tempo suficiente para mutações causarem a diferenciação de seus genomas, em média, em aproximadamente um a cada dois nucleotídios. Portanto, as sequências comuns aos genomas de camundongos e humanos provavelmente indicam funções comuns.

Os homólogos são identificados em virtude de apresentarem sequências de DNA similares. A análise do genoma de camundongo indica que o número de genes codificadores de proteínas que ele contém é semelhante ao do genoma humano. A inspeção adicional dos genes de camundongo revela que no mínimo 99% de todos os genes de camundongo apresentam algum homólogo no genoma humano e que no mínimo 99% de todos os genes humanos apresentam algum homólogo no genoma de camundongo. Portanto, os tipos de proteínas codificadas em cada genoma são essencialmente os mesmos. Além disso, aproximadamente 80% de todos os genes de camundongo e humanos são ortólogos claramente

identificáveis.

As semelhanças entre os genomas se estendem bem além do inventário dos genes codificadores de proteínas para a organização geral do genoma. Mais de 90% dos genomas humano e de camundongo podem ser separados em regiões correspondentes de **sintenia** conservada, nas quais a ordem dos genes dentro de blocos de tamanhos variados é a mesma da ordem no ancestral comum mais recente das duas espécies. Essa sintenia é muito útil para relacionar os mapas dos dois genomas. Por exemplo, o cromossomo humano 17 é ortólogo de um único cromossomo de camundongo (cromossomo 11). Embora tenha havido extensivos rearranjos intracromossômicos no cromossomo humano, existem 23 segmentos de sequências colineares com tamanho superior a 100 kb (Figura 14.17).

> CONCEITO-CHAVE Os genomas humano e de camundongo contêm conjuntos de genes semelhantes, com frequência dispostos em ordem semelhante.

Existem algumas diferenças detectáveis entre os inventários dos genes de camundongo e humanos. Em uma família de genes envolvidos na visão de cores, as opsinas, os seres humanos possuem um parálogo adicional. A presença dessa opsina equipou os seres humanos com a assim denominada visão tricromática, de modo que conseguimos perceber as cores ao longo de todo o espectro de luz visível — violeta, azul, verde, vermelho — enquanto o camundongo não consegue. Mas, novamente, a presença desse parálogo adicional em seres humanos e sua ausência em camundongos não nos informam, isoladamente, se ele foi adquirido na linhagem humana ou perdido na linhagem do camundongo. A análise dos genomas de outros primatas e mamíferos revelou que os primatas do Velho Mundo, tais como chimpanzés, gorilas e o macaco colobo, possuem esse gene, mas em todos os mamíferos não primatas ele está ausente. Podemos inferir com segurança a partir dessa distribuição filogenética do gene da opsina adicional que ele evoluiu em um ancestral dos primatas do Velho Mundo (o que inclui os seres humanos).

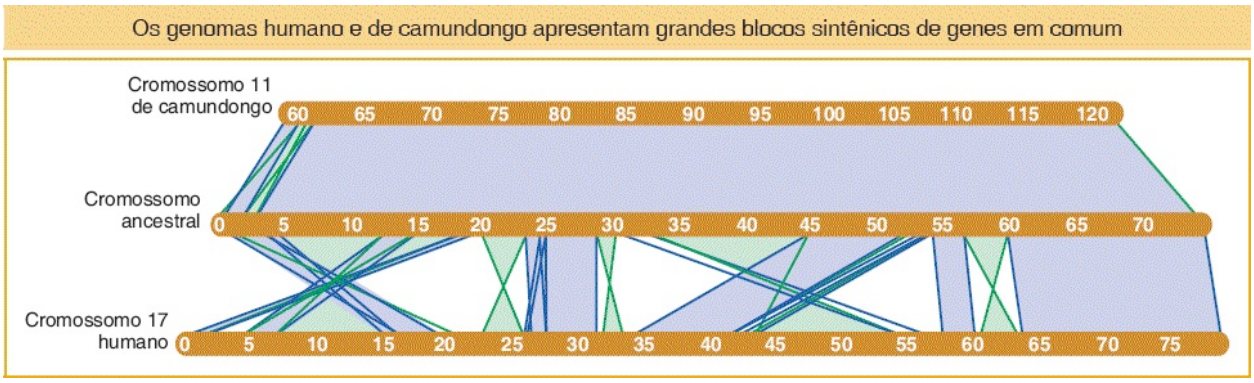


FIGURA 14.17 Sintenia entre o cromossomo 17 humano e o cromossomo 11 de camundongo. Grandes blocos sintênicos conservados com o tamanho de 100 kb ou mais estão demonstrados no cromossomo 17 humano, no cromossomo 11 de camundongo e no cromossomo inferido de seu último ancestral comum (reconstruído por meio da análise de genomas de outros mamíferos). Blocos diretos de sintenia estão demonstrados em *roxo-claro*; os blocos invertidos estão demonstrados em *verde*. Os tamanhos dos cromossomos estão indicados em megabases (Mb). (Dados de M. C. Zody *et al.*, "DNA Sequence of Human Chromosome 17 and Analysis of Rearrangement in the Human Lineage", *Nature* 440, 2006, 1045-1049, Fig. 2.)

Por outro lado, o genoma do camundongo contém mais cópias funcionais de alguns genes que refletem o seu estilo de vida. Camundongos apresentam aproximadamente 1.400 genes envolvidos no olfato — essa é a maior categoria funcional única de genes em seu genoma. Cães também apresentam uma grande quantidade de genes olfatórios. Isso certamente faz sentido para os estilos de vida das espécies. Camundongos e cães dependem fortemente do seu sentido de olfato e encontram odores diferentes daqueles notados por seres humanos. E o conjunto de genes olfatórios humanos, em comparação àquele de camundongos e cães, é surpreendentemente inferior. Apresentamos muitos genes olfatórios, mas uma grande fração deles é de pseudogenes com mutações inativadoras. Por exemplo, em apenas uma classe de genes olfatórios denominados genes *V1r*, camundongos apresentam aproximadamente 160 genes funcionais, mas apenas cinco dos aproximadamente 200 genes *V1r* no genoma humano são funcionais.

Ainda assim, essas diferenças no conteúdo gênico são relativamente modestas à luz das vastas diferenças na anatomia e no comportamento. A semelhança geral nos genomas de camundongos e seres humanos corresponde ao quadro que obtemos a partir do exame da *toolkit* genética que controla o desenvolvimento em diferentes espécies (ver Capítulo 13) — que grandes diferenças podem evoluir de

genomas que contêm conjuntos de genes semelhantes. Esse mesmo tema é ilustrado pela comparação de nosso genoma com aquele de nosso parente vivo mais próximo, o chimpanzé.

Genômica comparativa de chimpanzés e seres humanos

Os chimpanzés e os seres humanos apresentaram um último ancestral comum há aproximadamente 5 a 6 milhões de anos. Desde aquela época, as diferenças genéticas foram acumuladas por mutações que ocorreram em cada linhagem. O sequenciamento do genoma revelou que existem aproximadamente 35 milhões de diferenças de nucleotídios únicos entre chimpanzés e seres humanos, correspondendo a um grau de divergência de aproximadamente 1,06%. Além disso, aproximadamente 5 milhões de inserções e deleções, que variam em comprimento de apenas um único nucleotídio até mais de 15 kb, contribuem para um total de aproximadamente 90 Mb de sequência de DNA divergente (aproximadamente 3% do genoma total). A maior parte dessas inserções ou deleções está localizada fora das regiões codificadoras.

Em geral, as proteínas codificadas pelos genomas humano e de chimpanzé são extremamente semelhantes. Vinte e nove por cento de todas as proteínas ortólogas são *idênticas* em sequência. A maior parte das proteínas que diferem fazem isso por meio de apenas aproximadamente duas substituições de aminoácidos. Existem algumas diferenças detectáveis entre os chimpanzés e os seres humanos nos conjuntos de genes funcionais. Aproximadamente 80 genes que eram funcionais em seu ancestral comum deixaram de ser funcionais em seres humanos, em virtude de sua deleção ou do acúmulo de mutações. Algumas dessas alterações podem contribuir para diferenças na fisiologia.

Além das alterações em determinados genes, duplicações de segmentos cromossômicos em uma única linhagem contribuíram para a divergência genômica. Mais de 170 genes no genoma humano e mais de 90 genes no genoma de chimpanzé estão presentes em grandes segmentos duplicados. Essas duplicações são responsáveis por uma fração maior da divergência genômica total do que todas as mutações de nucleotídio único combinadas. Entretanto, ainda não está claro se elas contribuem para diferenças fenotípicas importantes.

É claro que todas as diferenças genéticas entre as espécies têm origem como variações dentro das espécies. O sequenciamento do genoma humano e o advento de métodos de sequenciamento de alta produtividade mais rápidos e menos dispendiosos abriram as portas para a análise detalhada da variação genética humana.

14.6 Genômica comparativa e medicina de seres humanos

A espécie humana, *Homo sapiens*, teve origem na África há aproximadamente 200.000 anos. Há cerca de 60.000 anos, populações deixaram a África e migraram por todo o mundo, povoando os cinco continentes. Essas populações migratórias encontraram diferentes climas, adotaram dietas diferentes e combateram diferentes patógenos em diferentes partes do mundo. Uma grande parte da história evolutiva recente da nossa espécie está registrada em nossos genomas, assim como as diferenças genéticas que tornam os indivíduos ou as populações mais ou menos suscetíveis a doenças.

Em geral, quaisquer dois genomas humanos não relacionados são 99,9% idênticos. A diferença de apenas 0,1% corresponde a aproximadamente 3 milhões de bases. Atualmente, o desafio é decifrar quais daquelas diferenças de bases são significativas em relação à fisiologia, ao desenvolvimento e à doença.

Após o avanço da sequência do primeiro genoma humano, tal conquista abriu as portas para uma análise muito mais rápida e menos dispendiosa de outros indivíduos. O motivo é que, com uma montagem de genoma conhecida como uma referência, é muito mais fácil alinhar as leituras de sequências brutas de indivíduos adicionais e desenhar abordagens para estudo e comparação de partes do genoma.

Uma das primeiras e maiores surpresas surgidas a partir da comparação de genomas humanos individuais é que os seres humanos diferem não meramente em uma base dentre um milhão, mas também no número de cópias de partes de genes individuais, genomas inteiros ou conjuntos de genes. Essas **variações no número de cópias (CNV)** incluem repetições e duplicações que aumentam o número de

cópias e deleções que reduzem o número de cópias. Entre quaisquer dois indivíduos não relacionados, pode haver 1.000 ou mais segmentos de DNA de comprimento superior a 500 pb que diferem no número de cópias. Algumas CNV podem ser razoavelmente grandes e abranger mais de 1 milhão de pares de bases.

Como tais números de cópias podem desempenhar um papel na evolução e nas doenças de seres humanos é de intenso interesse. Um caso no qual um aumento do número de cópias aparenta ter sido adaptativo diz respeito à dieta. Pessoas com dietas com alto teor de amido contêm, em média, mais cópias de um gene da amilase salivar (uma enzima que fragmenta o amido) do que pessoas com dietas tradicionalmente com baixo teor de amido. Em outros casos, variações no número de cópias foram associadas a síndromes, como o autismo.

Exoma e genômica personalizada

O QUE OS GENETICISTAS FAZEM ATUALMENTE

Avanços nas tecnologias de sequenciamento reduziram o custo do sequenciamento de genomas individuais de aproximadamente US$ 300 milhões em 2000, para US$ 1 milhão em 2008, para aproximadamente US$ 5.000 em 2013. Mas para muitos estudos em grande escala, esse ainda é um valor proibitivo. Para algumas aplicações, é mais prático e de custo eficaz, podendo ser tão informativo quanto, sequenciar apenas uma parte do genoma. Por exemplo, tendo em vista que muitas mutações que causam doenças ocorrem em sequências codificadoras, foram desenhadas estratégias para sequenciar todos os éxons ou o “exoma”, de indivíduos, assim como foi realizado no caso de Nicholas Volker.

A estratégia para o sequenciamento do exoma envolve a geração de uma biblioteca de DNA genômico que é enriquecida com sequências de éxons (Figura 14.18). O DNA é preparado por meio de (1) fragmentação do DNA genômico em segmentos unifilamentares curtos, (2) hibridização dos segmentos unifilamentares com sondas marcadas com biotina complementares às regiões dos éxons e purificação dos duplexes marcados com biotina, (3) amplificação dos duplexes ricos em éxons e (4) sequenciamento dos duplexes ricos em éxons. Desse modo, 30 a 60 megabases do genoma humano são alvo para o sequenciamento,

contrariamente às 3.200 megabases da sequência total.

Até o final de 2013, os exomas de mais de 100.000 indivíduos haviam sido sequenciados, no custo atual de apenas aproximadamente US$ 500 por exoma. Um poder do sequenciamento do exoma particularmente importante é a identificação das mutações *de novo* (mutações que não estão presentes em seus genitores) nos indivíduos. As referidas mutações são responsáveis por muitas doenças genéticas de aparecimento espontâneo cujas origens não seriam reveladas por meio de estudos tradicionais com base no heredograma. Como tal, o sequenciamento do exoma inteiro atualmente é uma ferramenta diagnóstica clínica em rápida difusão.

E assim como o sequenciamento do exoma pode ser utilizado para identificar diferenças genéticas entre os indivíduos, ele também pode ser utilizado para identificar diferenças entre células normais e anormais, tais como células cancerosas. O câncer é um grupo de doenças genéticas nas quais combinações de mutações de genes contribuem tipicamente para a perda do controle do crescimento e metástase. Compreender quais alterações genéticas são comuns a cânceres em particular ou a subconjuntos de cânceres não apenas aumentará a nossa compreensão sobre o câncer, mas também promete impactar o diagnóstico e o tratamento de modo poderoso. Pesquisadores em todo o mundo estão colaborando para criar um "atlas" de genomas de cânceres que compile nosso conhecimento em expansão sobre as mutações genéticas associadas a muitos cânceres (ver http://cancergenome.nih.gov para mais informações).

A capacidade de analisar rapidamente os genomas dos organismos também está impactando outras dimensões da medicina. Veremos um caso em seguida.

Genômica comparativa de E. coli não patogênica e patogênica

A *Escherichia coli* é encontrada em nossas bocas e em nossos tratos intestinais em grande número e essa espécie em geral é um simbionte benigno. Em virtude de seu papel central nas pesquisas genéticas, ela foi um dos primeiros genomas bacterianos sequenciados. O genoma de *E. coli* apresenta o tamanho de aproximadamente 4,6 Mb e contém 4.405 genes. Entretanto, realmente não é acurado denominá-lo "o genoma de *E. coli*". O primeiro genoma sequenciado foi derivado da linhagem de *E. coli* comum de laboratório K-12. Existem muitas

outras linhagens de *E. coli*, incluindo diversas importantes para a saúde humana.

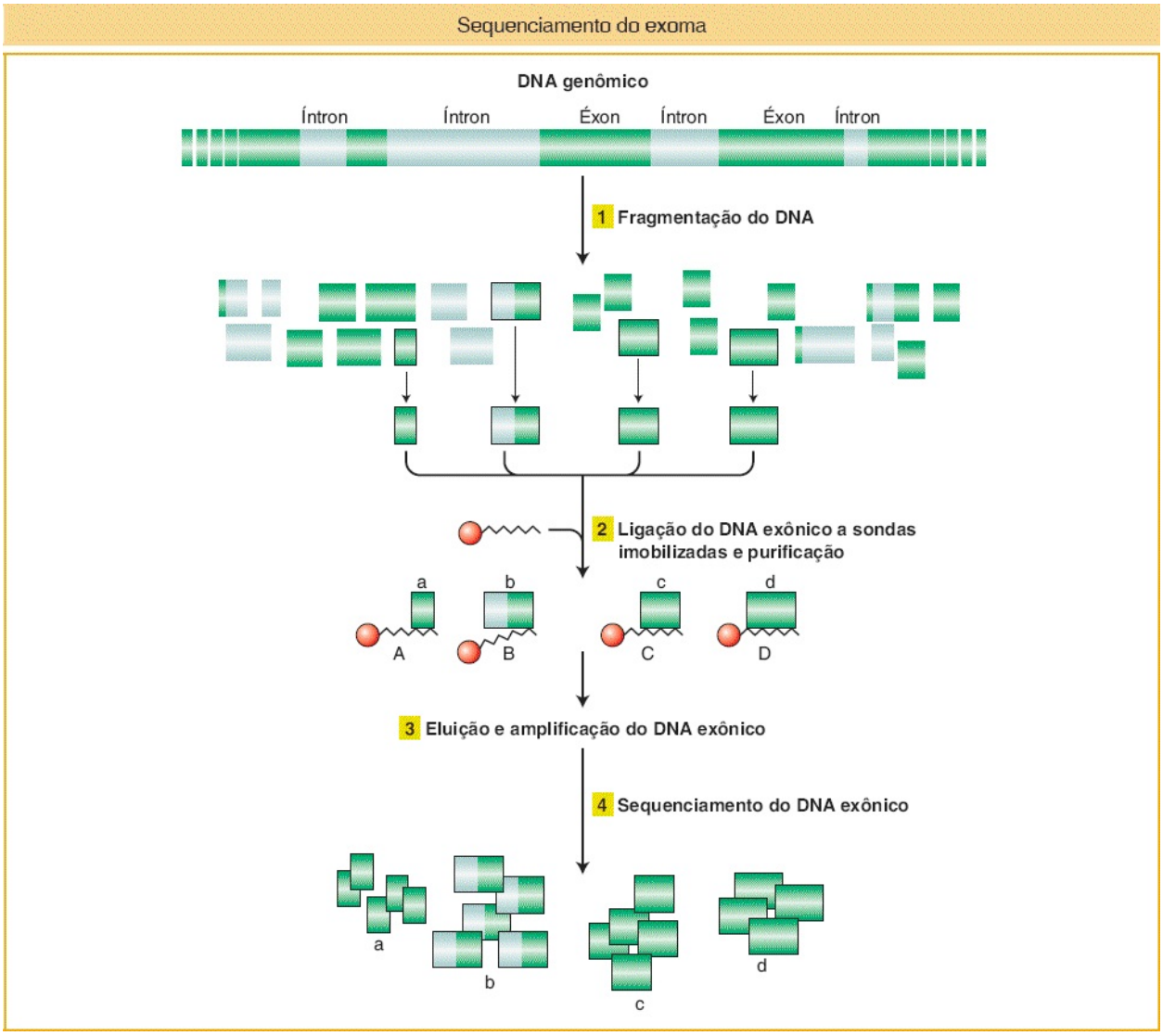


FIGURA 14.18 Com a finalidade de sequenciar apenas a fração exônica do genoma, o DNA genômico é fragmentado e desnaturado e os fragmentos que contêm éxons são hibridizados com sondas marcadas com biotina. Os duplexes que contêm as sondas pareadas em seguida são purificados e preparados para o sequenciamento.

Em 1982, um surto de uma doença humana em diversos estados dos EUA foi relacionado ao consumo de carne moída malcozida. A linhagem de *E. coli* O157:H7 foi identificada como a responsável e desde então tem sido associada a uma diversidade de surtos de infecção em grande escala. De fato, existe uma estimativa de 75.000 casos de infecção por *E. coli* anualmente nos EUA. Embora

quase todas as pessoas se recuperem da infecção, uma fração desenvolve síndrome hemolítico-urêmica, uma doença renal de risco potencial à vida.

Para compreender as bases genéticas da patogenicidade, o genoma de uma linhagem O157:H7 de *E. coli* foi sequenciado. As linhagens O157 e K-12 apresentam um arcabouço de 3.574 genes codificadores de proteínas em comum e a identidade média dos nucleotídios entre os genes ortólogos é de 98,4%, comparável àquela dos ortólogos de seres humanos e chimpanzés. Aproximadamente 25% dos ortólogos de *E. coli* codificam proteínas idênticas, semelhantes aos 29% de ortólogos de seres humanos e chimpanzés.

Apesar das semelhanças em muitas proteínas, os genomas e os proteomas diferem enormemente no conteúdo. O genoma de *E. coli* O157 codifica 5.416 genes, enquanto o genoma de *E. coli* K-12 codifica 4.405 genes. O genoma de *E. coli* O157 contém 1.387 genes que não são observados no genoma de K-12 e o genoma de K-12 contém 528 genes não observados no genoma de O157. A comparação dos mapas genômicos revela que os arcabouços comuns às duas linhagens estão entremeados com ilhas de genes específicos de K-12 ou O157 (Figura 14.19).

Entre os 1.387 genes específicos de *E. coli* O157 estão muitos genes que se suspeita codificarem fatores de virulência, incluindo toxinas, proteínas de invasão celular, proteínas de adesão e sistemas de secreção de toxinas, bem como possíveis genes metabólicos que podem ser necessários para o transporte de nutrientes, a resistência a antibióticos e outras atividades que podem conferir a capacidade de sobreviver em diferentes hospedeiros. A maior parte desses genes não era conhecida antes do sequenciamento e não seria conhecida atualmente se os pesquisadores tivessem confiado apenas na *E. coli* K-12 como um guia para todas as *E. coli*.

O surpreendente nível de diversidade entre dois membros da mesma espécie demonstra quão dinâmica a evolução do genoma pode ser. Acredita-se que a maior parte dos novos genes em linhagens de *E. coli* tenha sido introduzida por meio da transferência horizontal de genomas de vírus e outras bactérias (ver Capítulo 5). As diferenças também podem surgir por deleção gênica. Outras *E. coli* patogênicas e espécies bacterianas também exibem muitas diferenças no

conteúdo gênico de suas primas não patogênicas. A identificação dos genes que podem contribuir diretamente para a patogenicidade abre novos caminhos para compreensão, prevenção e tratamento de doenças infecciosas.

14.7 Genômica funcional e genética reversa

Geneticistas têm estudado a expressão e as interações de produtos gênicos individuais durante as últimas décadas. Com o advento da genômica, temos uma oportunidade de expandir esses estudos até um nível global por meio da utilização de abordagens de genoma total para estudar a maior parte dos ou todos os produtos gênicos de modo sistemático e simultâneo e em espécies que não são modelos experimentais previamente estabelecidos. Essa abordagem global para o estudo da função, da expressão e da interação dos produtos gênicos é denominada **genômica funcional**.

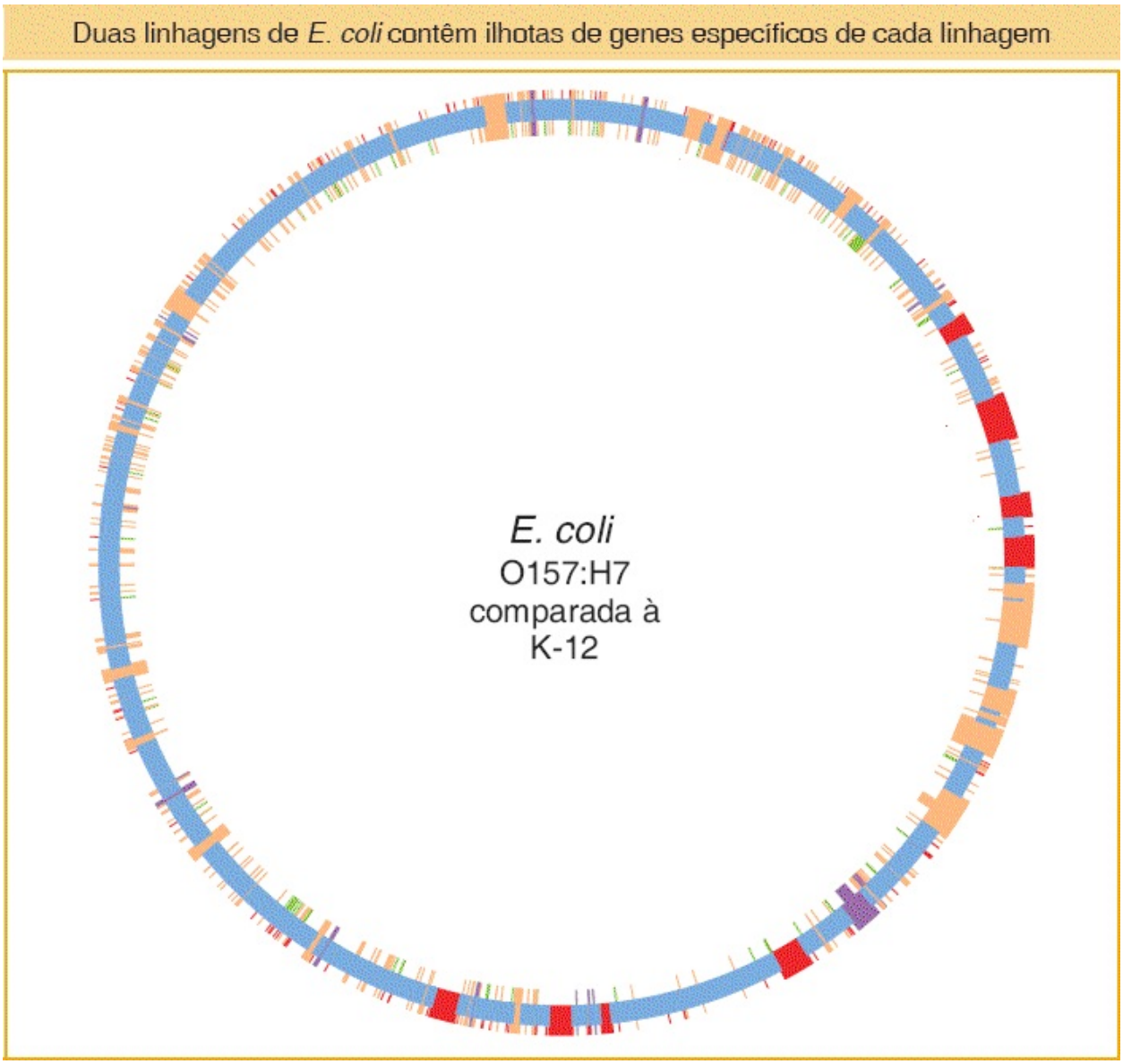


FIGURA 14.19 Os mapas genômicos circulares das linhagens de *E. coli* K-12 e O157:H7. O círculo ilustra a distribuição de sequências específicas de cada linhagem. O arcabouço colinear comum a ambas as linhagens está demonstrado em *azul*. As posições de sequências específicas de O157:H7 estão demonstradas em *vermelho*. As posições das sequências específicas de K-12 estão demonstradas em *verde*. As posições das sequências específicas de O157:H7 e K-12 no mesmo local estão demonstradas em *marrom*. As sequências hipervariáveis estão demonstradas em *roxo*. (Dados de N. T. Perna *et al.*, "Genome Sequence of Enterohaemorrhagic *Escherichia coli* O157:H7", *Nature* 409, 2001, 529-533. Cortesia de Guy Plunkett III e Frederick Blattner.)

"Omas"

Além do genoma, outros conjuntos de dados globais são de interesse. Seguindo o

exemplo do termo *genoma*, em relação ao qual "gene" mais "-oma" se torna uma palavra para "todos os genes", pesquisadores da genômica cunharam uma diversidade de termos para descrever outros conjuntos de dados globais com os quais estão trabalhando. Essa lista de *omas* inclui:

O transcriptoma. Sequências e padrões de expressão de todos os transcritos de RNA (quais tipos, onde nos tecidos, quando, quanto).

O proteoma. Sequências e padrões de expressão de todas as proteínas (onde, quando, quanto).

O interatoma. Conjunto completo de interações físicas de proteínas e segmentos de DNA, de proteínas e segmentos de RNA e entre proteínas.

Não consideraremos todos esses *omas* nesta seção, mas enfocaremos algumas das técnicas globais que estão começando a ser exploradas para obter esses conjuntos de dados.

Utilização de microarranjos de DNA para estudar o transcriptoma. Suponha que desejamos responder à questão: que genes são ativos em uma célula em particular sob determinadas condições? Essas condições podem ser um ou mais estágios no desenvolvimento ou podem ser a presença ou a ausência de um patógeno ou de um hormônio. Os genes ativos são transcritos em RNA e, assim, o conjunto de transcritos de RNA presentes na célula pode nos informar quais genes estão ativos. Aqui é onde a nova tecnologia de *chips* de DNA utilizada para analisar transcritos de DNA é tão poderosa.

Os *chips* de DNA são amostras de DNA dispostas como uma série de manchas microscópicas ligadas a um "*chip*" de vidro do tamanho de uma lamínula de microscópio. O conjunto de DNA assim demonstrado é denominado **microarranjo**. Um tipo de microarranjo típico contém oligonucleotídios sintéticos curtos que representam a maior parte dos genes, ou todos, em um genoma (Figura 14.20). Os microarranjos de DNA têm enriquecido a genética molecular ao possibilitar a análise de transcritos de RNA para todos os genes simultaneamente em um único experimento. Vejamos como atua esse processo em mais detalhes.

Microarranjos podem detectar diferenças na expressão gênica

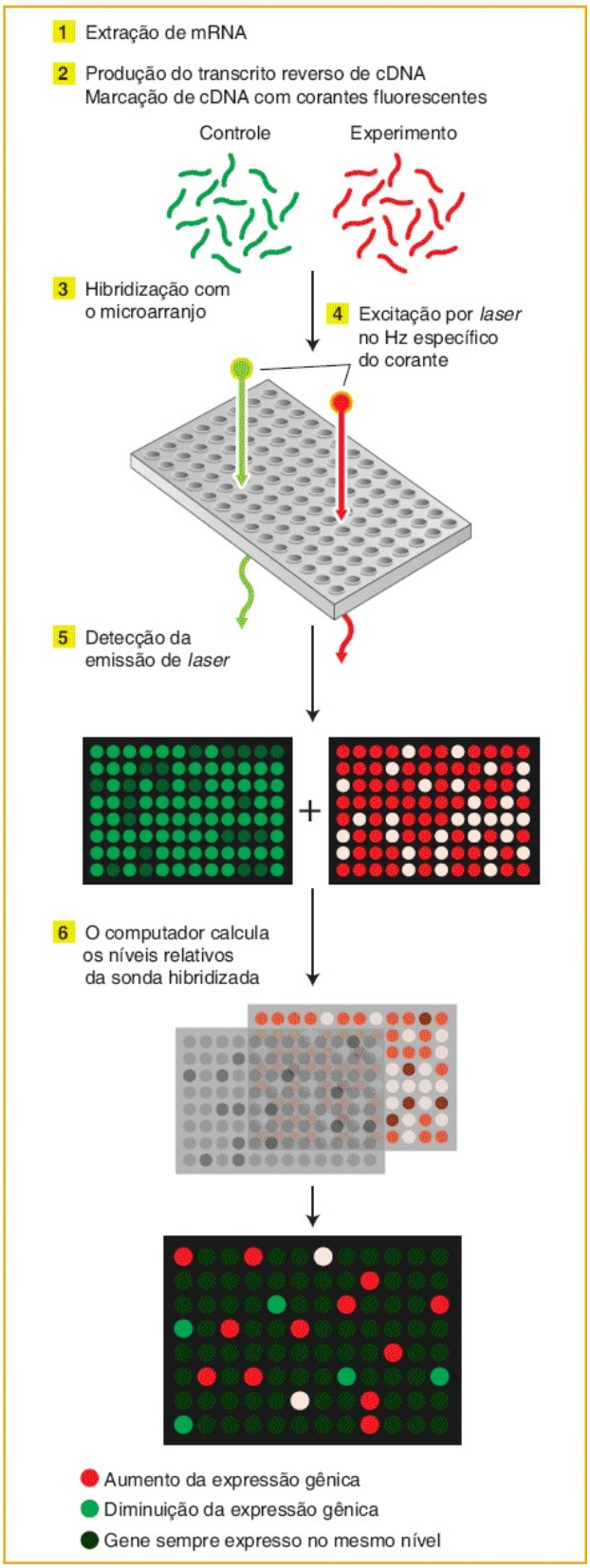


FIGURA 14.20 As principais etapas em uma análise com microarranjo são (1) extração do mRNA de células ou tecidos, (2) síntese de sondas de cDNA marcadas com corante fluorescente, (3) hibridização com o microarranjo, (4) detecção do sinal fluorescente de sondas hibridizadas e (5) análise da imagem para identificar os níveis relativos da sonda hibridizada. Os níveis relativos revelam os genes cuja expressão está aumentada ou diminuída sob as condições analisadas.

Os microarranjos são expostos a sondas de cDNA — por exemplo, um conjunto de sondas utilizadas como controle e um conjunto de sondas representando uma condição específica. O conjunto utilizado como um controle pode ser produzido a partir do conjunto total de moléculas de RNA extraídas de um tipo celular em particular cultivado sob condições típicas. O segundo conjunto de sondas pode ser produzido a partir do RNA extraído de células cultivadas em alguma condição experimental. Marcações fluorescentes são ligadas às sondas e as sondas são hibridizadas com o microarranjo. A ligação relativa das moléculas da sonda com o microarranjo é automaticamente monitorada com a utilização de um microscópio iluminado com feixe de *laser*. Desse modo, são identificados os genes cujos níveis de expressão estão aumentados ou diminuídos sob a determinada condição experimental. De modo semelhante, os genes que estão ativos em um determinado tipo celular ou em um determinado estágio do desenvolvimento podem ser identificados.

Com a compreensão sobre quais genes estão ativos ou inativos em um determinado estágio do desenvolvimento, em um tipo celular em particular ou em diversas condições ambientais, os conjuntos de genes que podem responder a sinais reguladores semelhantes podem ser identificados. Além disso, os perfis de expressão gênica podem pintar um quadro das diferenças entre as células normais e doentes. Ao identificar os genes cuja expressão é alterada por mutações, em células cancerosas ou por um patógeno, os pesquisadores podem planejar novas estratégias terapêuticas.

Utilização do teste di-híbrido para estudar o interatoma proteína-proteína. Uma das atividades mais importantes das proteínas é sua interação com outras proteínas. Em virtude do grande número de proteínas em qualquer célula, biólogos buscam modos para estudar sistematicamente todas as interações de

proteínas individuais em uma célula. Um dos modos mais comuns de estudar o interatoma utiliza um sistema criado em células de levedura denominado **teste di-híbrido**, que detecta interações físicas entre duas proteínas. A base do teste é o ativador transcricional codificado pelo gene *GAL4* de levedura (ver Capítulo 12).

Lembre que essa proteína apresenta dois domínios: (1) um domínio de ligação ao DNA que se liga ao sítio de início de transcrição e (2) um domínio de ativação que ativará a transcrição, mas que por si próprio não consegue se ligar ao DNA. Portanto, os dois domínios devem estar próximos para que ocorra a ativação da transcrição. Suponha que você está investigando se duas proteínas interagem. A estratégia do sistema di-híbrido é separar os dois domínios do ativador codificado por *GAL4*, tornando a ativação de um gene repórter impossível. Cada domínio é conectado a uma proteína diferente. Se as duas proteínas interagem, elas unirão os dois domínios. O ativador se tornará ativo e iniciará a transcrição do gene repórter.

Como esse esquema é implementado na prática? O gene *GAL4* é dividido entre dois plasmídios, de modo que um plasmídio contém a parte que codifica o domínio de ligação ao DNA e o outro plasmídio contém a parte que codifica o domínio de ativação. Em um plasmídio, um gene para uma proteína em investigação é recomposto próximo ao domínio de ligação ao DNA e essa proteína de fusão atua como uma "isca". No outro plasmídio, um gene para outra proteína em investigação é recomposto próximo ao domínio de ativação e diz-se que essa proteína de fusão é o "alvo" (Figura 14.21). Os dois plasmídios híbridos em seguida são introduzidos na mesma célula de levedura — talvez por meio do cruzamento de células haploides que contêm a isca e os plasmídios-alvo. A etapa final é procurar pela ativação da transcrição por um construto de gene repórter regulado por *GAL4*, que seria a comprovação de que aquela isca e o alvo ligaram-se. O sistema di-híbrido pode ser automatizado para possibilitar a busca por interações proteicas em todo o proteoma.

Estudo do interatoma proteína-DNA com a utilização de análise de imunoprecipitação de cromatina (ChIP). A ligação sequência-específica de proteínas ao DNA é crítica para a expressão gênica correta. Por exemplo,

proteínas reguladoras se ligam a promotores e ativam ou reprimem a transcrição tanto em bactérias quanto em eucariotos (ver Capítulos 11, 12 e 13). No caso dos eucariotos, os cromossomos estão organizados em cromatina, na qual a unidade fundamental, o nucleossomo, contém DNA enrolado ao redor de histonas. A modificação pós-tradução das histonas com frequência dita quais proteínas se ligam e onde (ver Capítulo 12). Foi desenvolvida uma diversidade de tecnologias que possibilitam que os pesquisadores isolem regiões específicas da cromatina, de modo que o DNA e suas proteínas correlatas possam ser analisados em conjunto. O método mais amplamente utilizado é denominado **ChIP** (em referência a **imunoprecipitação da cromatina**) e a sua aplicação está descrita a seguir (Figura 14.22).

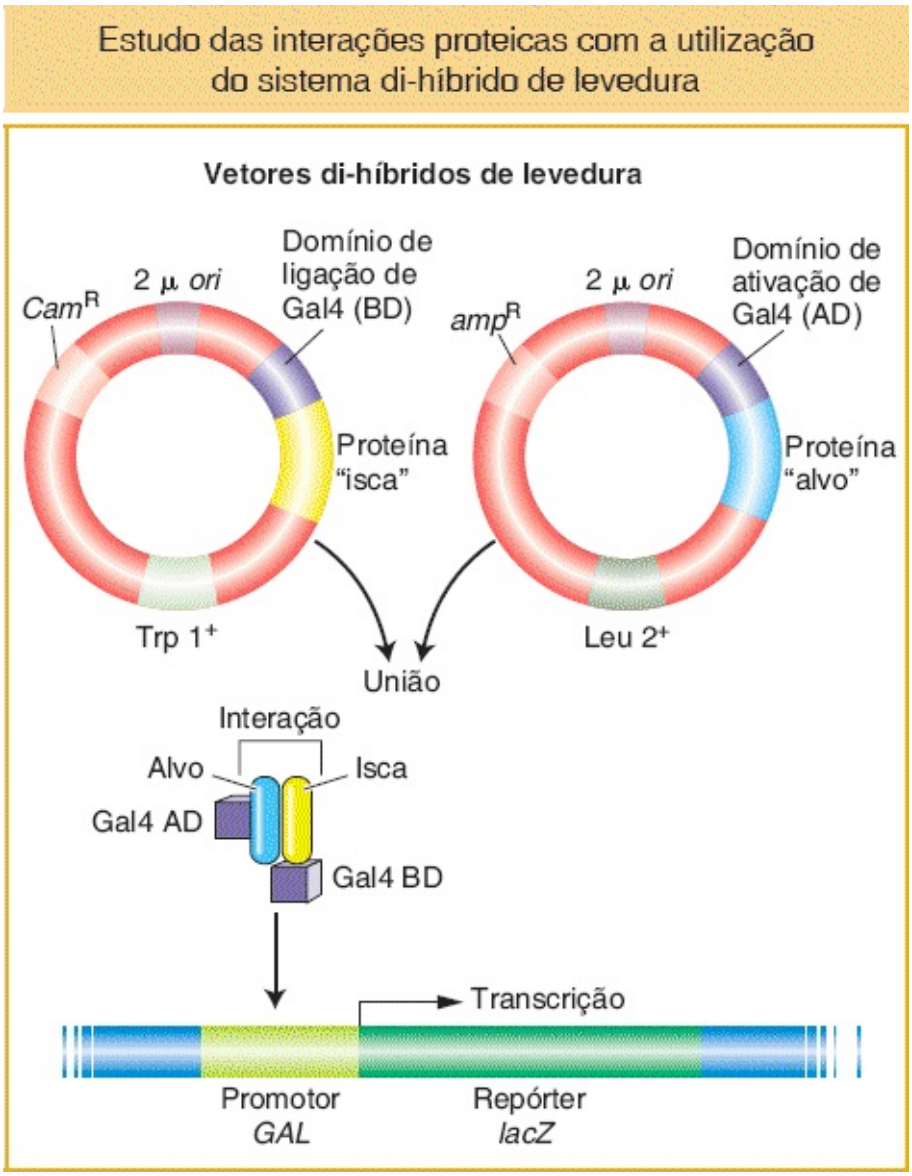


FIGURA 14.21 O sistema utiliza a ligação de duas proteínas, uma proteína "isca" e uma proteína "alvo", para restaurar a função da proteína Gal4, que ativa um gene repórter. *Cam*, Trp e Leu são componentes dos sistemas de seleção para a movimentação dos plasmídios entre as células. O gene repórter é *lacZ*, que está localizado em um cromossomo de levedura (*em azul*).

Digamos que você isolou um gene de levedura e suspeita que ele codifique uma proteína que se liga ao DNA quando a levedura é cultivada em alta temperatura. Você deseja saber se essa proteína se liga ao DNA e, caso afirmativo, a qual sequência na levedura. Um modo de abordar essa questão é primeiramente tratar as células de levedura que foram cultivadas em alta temperatura com uma

substância química que realizará a ligação cruzada das proteínas com o DNA. Desse modo, as proteínas ligadas ao DNA na ocasião do isolamento da cromatina permanecerão ligadas durante os tratamentos subsequentes. A próxima etapa é quebrar a cromatina em pequenos pedaços. Para separar o fragmento que contém seu complexo proteína-DNA de outros, você utiliza um anticorpo que reage especificamente com a proteína codificada. Você adiciona seu anticorpo à mistura, de modo a formar um complexo imune que pode ser purificado. O DNA ligado ao complexo imune pode ser analisado após a reversão da ligação cruzada. O DNA ligado à proteína pode ser sequenciado diretamente ou amplificado em muitas cópias por meio de PCR para preparar para o sequenciamento do DNA.

Conforme você viu no Capítulo 12, as proteínas reguladoras com frequência ativam a transcrição de muitos genes simultaneamente por meio da ligação a diversas regiões promotoras. Uma variação do procedimento de ChIP, denominada ChIP-chip, foi planejada para identificar múltiplos sítios de ligação em um genoma sequenciado. As proteínas que se ligam a muitas regiões genômicas são imunoprecipitadas conforme descrito anteriormente. Em seguida, após a reversão da ligação cruzada, os fragmentos de DNA são marcados e utilizados como sondas em *chips* de microarranjo que contêm a sequência genômica inteira da espécie em estudo.

Genética reversa

Os tipos de dados obtidos a partir de experimentos de microarranjos e triagens de interação proteica são sugestivos de interações no genoma e no proteoma, mas não possibilitam que sejam obtidas conclusões firmes a respeito das funções dos genes e das interações *in vivo*. Por exemplo, a descoberta de que a expressão de determinados genes é perdida em alguns cânceres não é uma comprovação de causa e efeito. O padrão-ouro para o estabelecimento da função de um gene ou de um elemento genético é romper a sua função e compreender os fenótipos em condições nativas. Com início a partir das sequências gênicas disponíveis, pesquisadores atualmente podem utilizar uma diversidade de métodos para romper a função de um gene específico. Esses métodos são denominados **genética reversa**. A análise por genética reversa tem início com uma molécula

conhecida — uma sequência de DNA, um mRNA ou uma proteína — e em seguida realizam-se tentativas para romper essa molécula para avaliar a função do produto gênico normal na biologia do organismo.

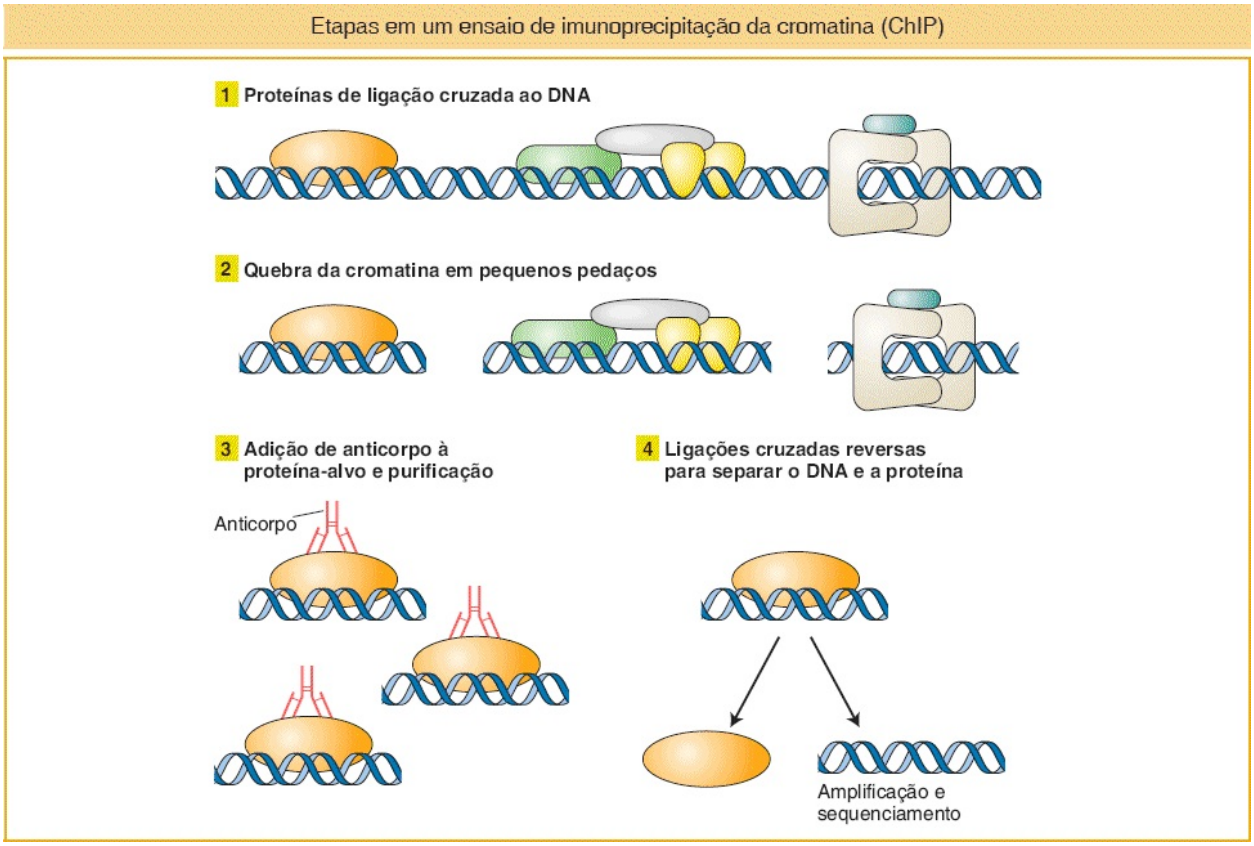


FIGURA 14.22 ChIP é uma técnica para o isolamento do DNA e suas proteínas correlatas em uma região específica da cromatina, de modo que ambos possam ser analisados em conjunto.

Existem diversas abordagens para a genética reversa e novas tecnologias estão sendo constantemente desenvolvidas e refinadas. Uma abordagem é introduzir mutações aleatórias no genoma, mas em seguida enfocar o gene de interesse por meio da identificação molecular de mutações no gene. Uma segunda abordagem é conduzir uma mutagênese direcionada que produza mutações diretamente no gene de interesse. Uma terceira abordagem é criar *fenocópias* — efeitos comparáveis a fenótipos mutantes — normalmente por meio de tratamento com agentes que interfiram com o transcrito de mRNA do gene.

Cada abordagem tem suas vantagens. A mutagênese aleatória está bem-

estabelecida, mas requer que todas as mutações sejam peneiradas para encontrar aquelas que incluem o gene de interesse. A mutagênese direcionada também pode ser de trabalho intenso, mas, após a obtenção da mutação-alvo, sua caracterização é mais direta. A criação de fenocópias pode ser muito eficiente, especialmente como bibliotecas de ferramentas que foram desenvolvidas para espécies-modelo em particular. Consideraremos exemplos de cada uma dessas abordagens.

Genética reversa por meio de mutagênese aleatória. A mutagênese aleatória para genética reversa emprega os mesmos tipos de mutágenos gerais que são utilizados para a genética direta: agentes químicos, radiação ou elementos genéticos de transposição (ver Capítulo 5). Entretanto, em vez de triar o genoma como um todo à procura de mutações que exerçam um efeito fenotípico em particular, a genética reversa enfoca o gene em questão, o que pode ser realizado por um de dois modos gerais.

Uma abordagem é enfocar a localização do gene no mapa. Apenas mutações que se encontram na região do genoma na qual o gene está localizado são retidas para a análise molecular detalhada adicional. Portanto, nessa abordagem, as mutações recuperadas devem ser mapeadas. Um modo direto é cruzar um novo mutante com um mutante que contenha uma deleção ou mutação conhecida do gene de interesse. Simbolicamente, o pareamento é *mutante novo/mutante conhecido*. Apenas os pareamentos que resultam em progênie com um fenótipo mutante (que demonstram ausência de complementação) são salvos para estudo.

Em outra abordagem, o gene de interesse é identificado no genoma mutagenizado e verificado em relação à presença de mutações. Por exemplo, se o mutágeno causa pequenas deleções, então, após a amplificação por PCR dos fragmentos gênicos, os genes dos genomas parental e mutagenizado podem ser comparados, na busca por um genoma mutagenizado no qual o gene de interesse seja de tamanho reduzido. De modo semelhante, inserções de elementos de transposição no gene de interesse podem ser prontamente detectadas em virtude do aumento de seu tamanho. Com a capacidade de sequenciar genomas inteiros de espécies-modelo rapidamente e com custo relativamente baixo, pode-se também buscar por substituições de um só par de bases. Desses modos, um conjunto de

genomas que contém mutações aleatórias pode ser efetivamente triado para identificar a pequena fração de mutações que são de interesse para o pesquisador.

Genética reversa por mutagênese direcionada. Durante a maior parte do século 20, pesquisadores consideraram a possibilidade de direcionar mutações para um gene específico como o "santo graal" inatingível da genética. Entretanto, atualmente diversas das técnicas estão disponíveis. Após a inativação de um gene em um indivíduo, os geneticistas podem avaliar o fenótipo exibido em busca de pistas da função do gene. Embora as ferramentas para as mutações gênicas direcionadas tenham sido desenvolvidas pela primeira vez com a utilização de técnicas para organismos-modelo, novas tecnologias estão possibilitando a ruptura e a manipulação de genes em espécies não modelo.

A mutagênese de gene específico normalmente requer a substituição de uma cópia do tipo selvagem residente de todo um gene por uma versão mutada desse gene. O gene mutado é inserido no cromossomo por meio de um mecanismo que se assemelha à recombinação homóloga, substituindo a sequência normal pela mutante (Figura 14.23). Essa abordagem pode ser utilizada para o nocaute de um gene-alvo, no qual um alelo nulo substitui a cópia do tipo selvagem. Algumas técnicas são tão eficientes que, em *E. coli* e *S. cerevisiae*, por exemplo, foi possível produzir mutações em todos os genes do genoma para tentar determinar sua função biológica.

> CONCEITO-CHAVE A mutagênese direcionada é o meio mais preciso para obter mutações em um gene específico e atualmente pode ser praticada em uma diversidade de sistemas-modelo, incluindo camundongos e moscas.

Genética reversa por meio da criação de fenocópias. A vantagem da inativação de um gene é que as mutações serão transmitidas de uma geração para a próxima e, assim, após sua obtenção, uma linhagem de mutantes estará sempre disponível para futuros estudos. Por outro lado, a criação de fenocópias pode ser aplicada a muitos organismos, independentemente de quão desenvolvida é a tecnologia

genética para uma determinada espécie.

Uma das descobertas mais animadoras da última década foi a de um mecanismo difundido cuja função natural aparenta ser proteger a célula de um DNA exógeno. Esse mecanismo é denominado **RNA de interferência (RNAi)**, descrito no Capítulo 8. Pesquisadores tiraram vantagem desse mecanismo celular a fim de desenvolver um método poderoso para inativar genes específicos. A inativação é conquistada como segue. Um RNA bifilamentar é produzido com sequências homólogas à parte do gene em estudo e é introduzido em uma célula (Figura 14.24). Em seguida, o complexo de silenciamento induzido pelo RNA ou RISC degrada o mRNA nativo que é complementar ao RNA bifilamentar. O resultado é uma redução completa ou considerável dos níveis de mRNA que dura horas ou dias, anulando, assim, a expressão daquele gene. A técnica tem sido amplamente aplicada em sistemas-modelo tais como *C. elegans*, *Drosophila*, peixe-zebra e diversas espécies de plantas.

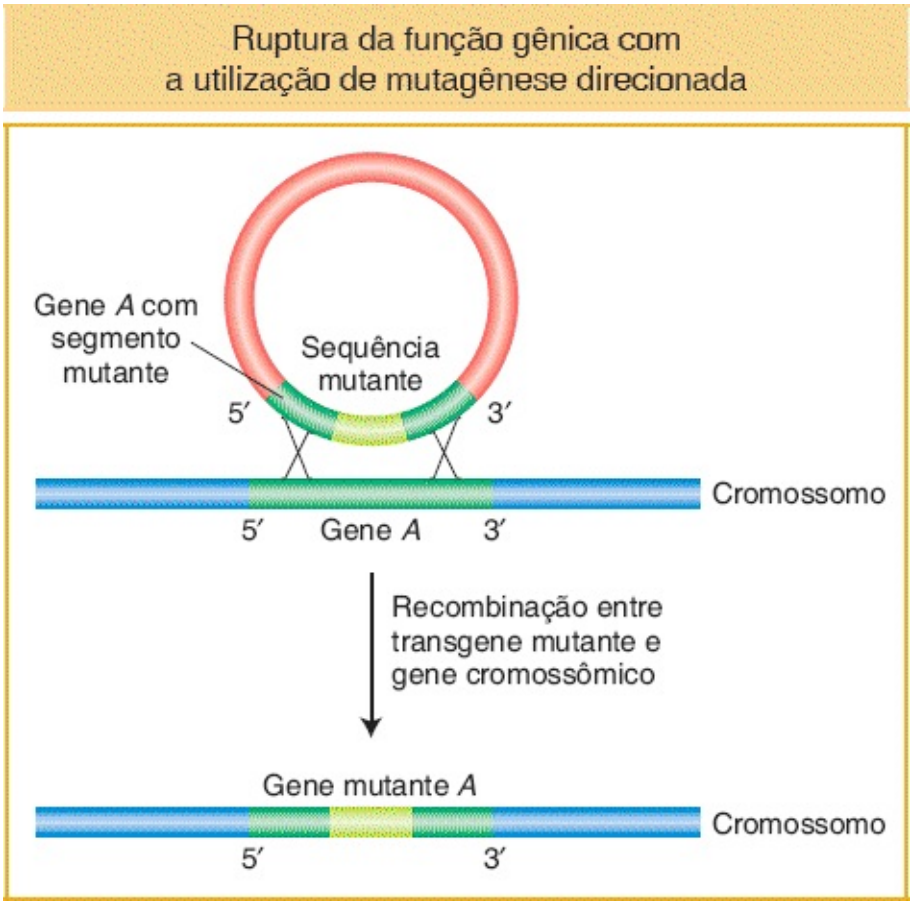


FIGURA 14.23 Evento molecular básico na substituição gênica direcionada. Um transgene que contém sequências de duas extremidades de um gene, mas com um segmento de DNA selecionável no meio, é introduzido em uma célula. A dupla recombinação entre o transgene e um gene cromossômico normal produz um gene cromossômico recombinante que incorporou o segmento anormal.

Mas o que torna o RNAi especialmente poderoso é que ele pode ser aplicado a organismos não modelo. Primeiramente, os genes-alvo de interesse podem ser identificados por meio de genômica comparativa. Em seguida, são produzidas sequências de RNAi para promover a inibição de genes-alvo específicos. Essa técnica tem sido aplicada, por exemplo, para um mosquito de transmite malária (*Anopheles gambiae*). Com a utilização dessas técnicas, os cientistas conseguem compreender melhor os mecanismos biológicos relacionados com o efeito clínico ou econômico da referida espécie. Os genes que controlam o complicado ciclo de vida do parasita da malária, parcialmente no mosquito hospedeiro e parcialmente no corpo humano, podem ser mais bem-compreendidos, revelando novos modos de controlar a doença infecciosa mais comum no mundo.

> CONCEITO-CHAVE Os métodos baseados em RNAi proporcionam modos gerais de interferir experimentalmente com a função de um gene específico sem alterar a sua sequência de DNA (em geral denominado *fenocópia*).

Genômica funcional com organismos não modelo. Uma grande parte de nossa consideração da dissecção mutacional e da criação de fenocópias enfocou em organismos-modelo genéticos. Um atual foco de muitos geneticistas é a aplicação mais ampla dessas técnicas, incluindo para espécies que apresentam efeitos negativos sobre a sociedade humana, tais como parasitas, transmissores de doenças ou pragas agrícolas. As técnicas genéticas clássicas não são prontamente aplicáveis para maior parte dessas espécies, mas os papéis de genes específicos podem ser avaliados por meio da inserção de transgenes e da criação de fenocópias.

A inserção de transgenes em um besouro está demonstrada na Figura 14.25. Besouros transgênicos podem ser produzidos por meio da utilização de

metodologia semelhante àquela utilizada para produzir *Drosophila* transgênica (ver Capítulo 10). Entretanto, é necessário algum modo para identificar a transgênese de sucesso. Portanto, a técnica depende da utilização de um gene repórter que possa ser expresso em um receptor do tipo selvagem. A proteína fluorescente verde (GFP), originalmente isolada a partir de água-viva, é um repórter útil para essa aplicação. Assim como em *Drosophila*, os transgenes são inseridos como partes de transpósons e um plasmídio auxiliar que codifica uma transposase facilita a inserção do transpóson que contém o transgene. A Figura 14.25 demonstra a utilização de transgenes GFP direcionados por um elemento acentuador que orienta a expressão no olho do inseto. Esse método também foi utilizado efetivamente para criar transgenes que expressam GFP em espécies de mosquito que transmitem a febre amarela e a dengue (*Aedes aegypti*), em um besouro da farinha (*Tribolium castaneum*) e no bicho-da-seda (*Bombyx mori*) (Figura 14.26). Com frequência, o transgene GFP é utilizado simplesmente como um marcador genético para experimentos nos quais o construto de RNAi ou outro transgene é coinserido com a finalidade de manipular a função gênica.

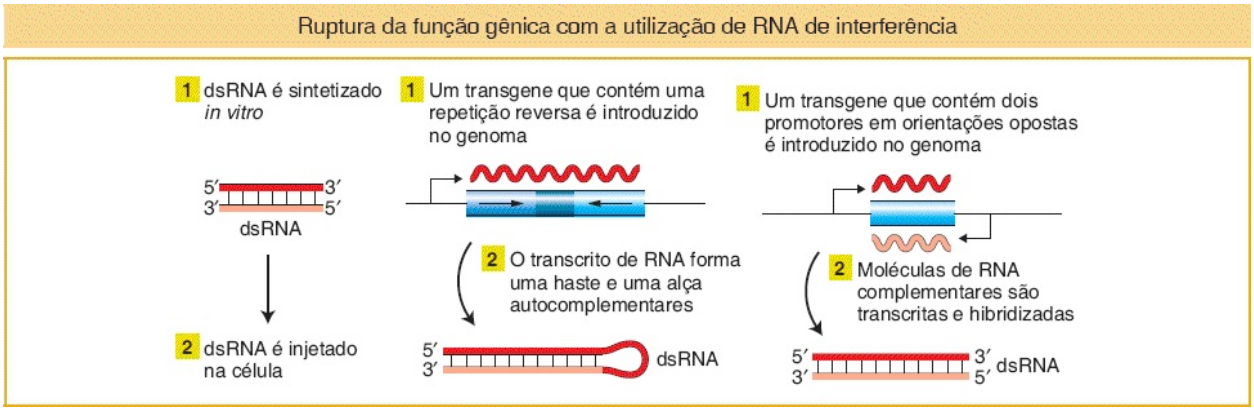


FIGURA 14.24 Três modos para criar e introduzir um RNA bifilamentar (dsRNA) em uma célula. O dsRNA em seguida estimulará a RNAi, degradando sequências que correspondem àquelas no dsRNA.

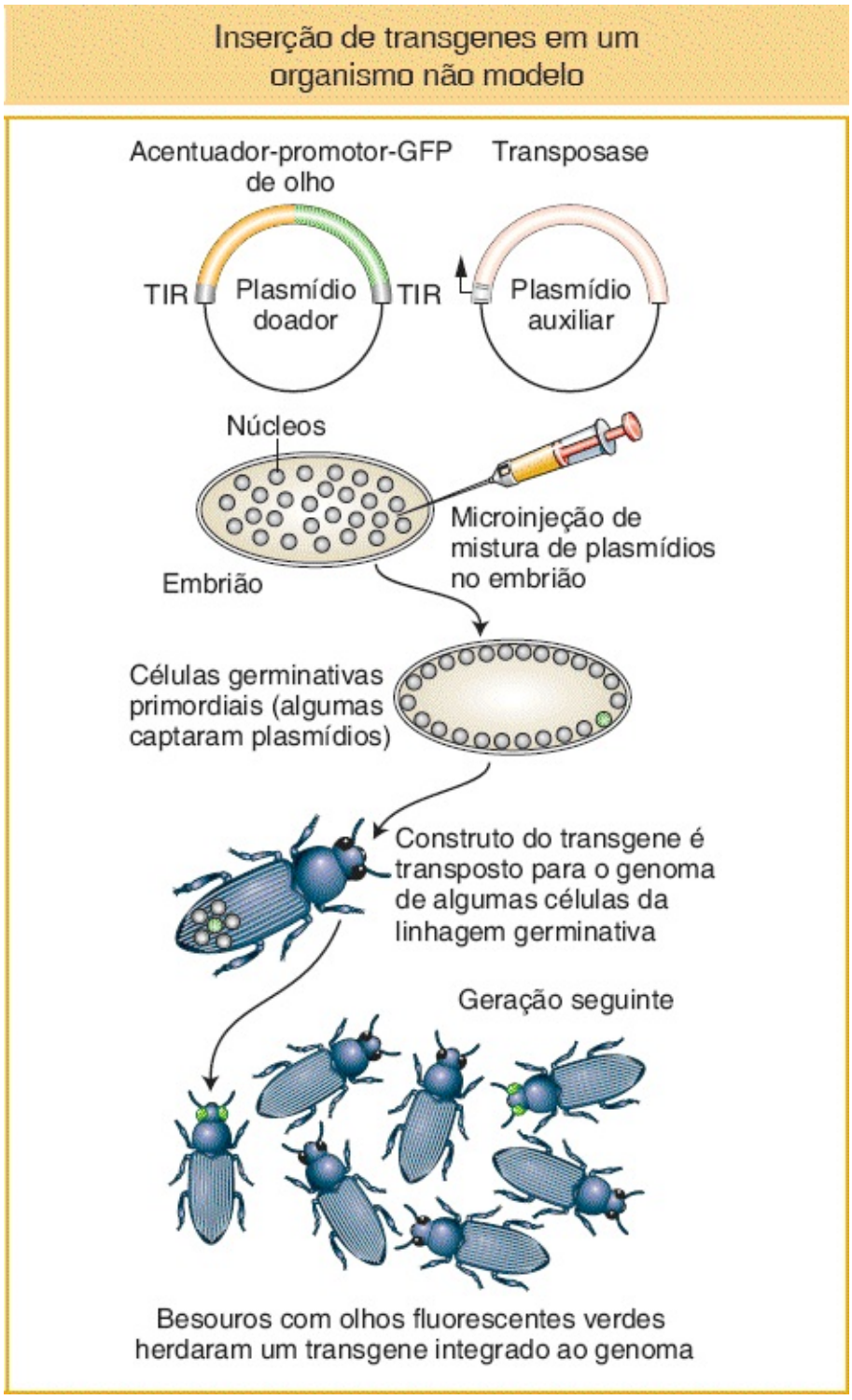


FIGURA 14.25 Criação de besouros transgênicos que expressam uma proteína fluorescente verde. TIR = Repetição invertida terminal.

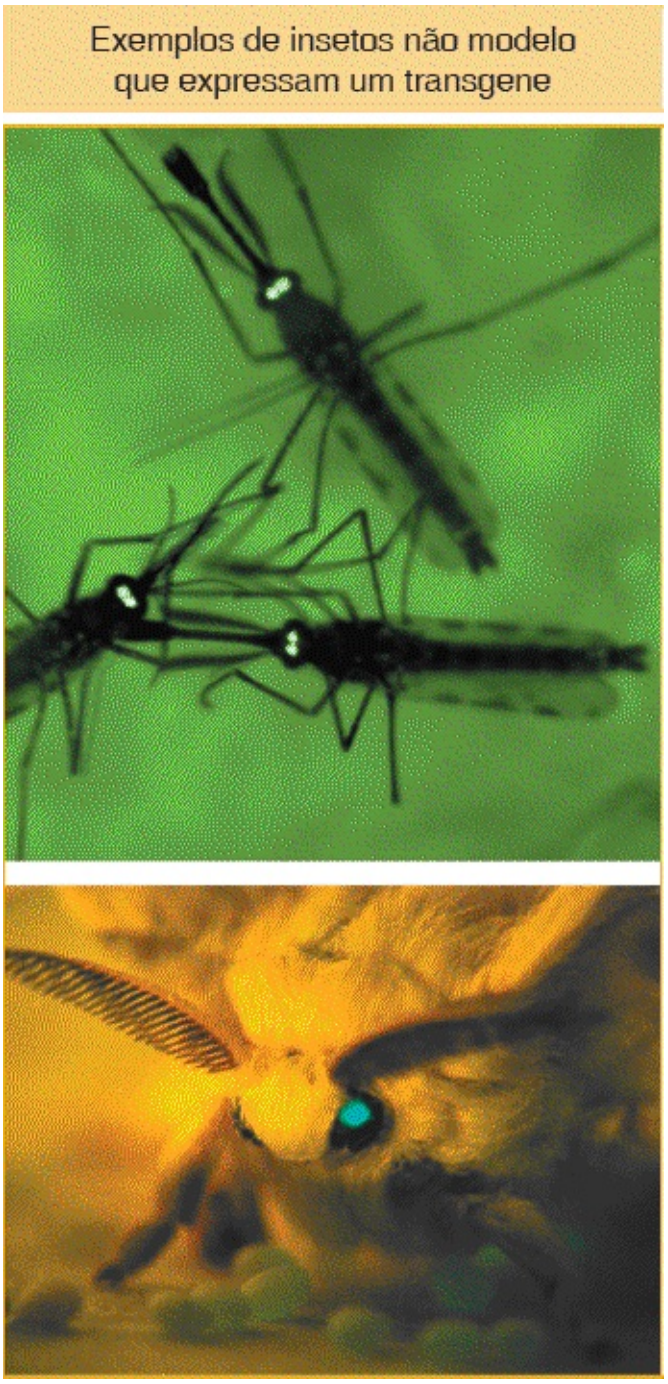


FIGURA 14.26 Exemplos de uma proteína repórter fluorescente verde transgênica expressa nos olhos de alguns insetos não modelo. A expressão é ativada a partir de um único promotor ativo nos olhos. Os insetos são o mosquito (*Aedes aegypti*, *parte superior*) e o bicho-da-seda (*Bombyx mori*, *parte inferior*). (*Parte superior*, AP Photo/Jacquelyn Martin; *parte inferior*, Marek Jindra.)

RESUMO

A análise genômica adota as abordagens da análise genética e as aplica à coleção de conjuntos de dados globais para conquistar objetivos tais como o mapeamento e o sequenciamento de genomas inteiros e a caracterização de todos os transcritos e de todas as proteínas. As técnicas genômicas requerem o rápido processamento de grandes conjuntos de material experimental, todas dependentes de extensiva automação.

O problema-chave na compilação de uma sequência precisa de um genoma é obter leituras de sequências curtas e relacioná-las entre si por meio da identidade de sequência para construir uma sequência de consenso de um genoma inteiro. Isso pode ser realizado diretamente em relação aos genomas bacterianos ou de ancestrais por meio do alinhamento de sequências sobrepostas de diferentes leituras de sequência para compilar o genoma inteiro, tendo em vista que existem poucos segmentos de DNA, ou nenhum, presentes em mais de uma cópia nos referidos organismos. O problema é que os genomas complexos de plantas e animais estão repletos de tais sequências repetitivas. Essas sequências repetitivas interferem com a produção de sequências contigs precisas. O problema é resolvido pelo sequenciamento *shotgun* de genoma inteiro (WGS) com a utilização de leituras de pontas pareadas.

A apresentação de um mapa de sequência genômica fornece o texto criptografado e bruto do genoma. O papel da bioinformática é interpretar essa informação criptografada. Para a análise dos produtos gênicos, são utilizadas técnicas de informática para identificar ORF e RNA não codificadores, em seguida integrar esses resultados com as evidências experimentais disponíveis para estruturas transcritas (sequências de cDNA), semelhanças proteicas e conhecimento de *motifs* de sequências características.

Um dos meios mais poderosos para o avanço da análise e da anotação dos genomas é por meio da comparação dos genomas de espécies relacionadas. A conservação de sequências entre espécies é um guia confiável para a identificação de sequências funcionais nos genomas complexos de muitos animais e plantas. A genômica comparativa também pode revelar como os genomas foram alterados no curso da evolução e como essas alterações podem estar relacionadas com diferenças na fisiologia, na anatomia ou no comportamento das espécies. As

comparações dos genomas humanos estão acelerando a descoberta de mutações de doenças raras. Na genômica bacteriana, comparações de linhagens patogênicas e não patogênicas revelaram muitas diferenças no conteúdo gênico que podem contribuir para a patogenicidade.

A genômica funcional tenta compreender o funcionamento do genoma como um todo. Dois elementos-chave são o transcriptoma, o conjunto de todos os transcriptos produzidos, e o interatoma, o conjunto de produtos gênicos que interagem e outras moléculas que em conjunto possibilitam que uma célula seja produzida e funcione. A função dos genes individuais e dos produtos gênicos em relação aos quais mutações clássicas não se encontram disponíveis pode ser testada por meio de genética reversa — por mutação direcionada ou criação de fenocópias.

TERMOS-CHAVE

anotação
arcabouço
biblioteca de DNA molde
biblioteca genômica
bioinformática
ChIP (análise de imunoprecipitação da cromatina)
etiqueta de sequência expressa (EST)
exoma
filogenia
genética reversa
genômica
genômica comparativa
genômica funcional
grupo evolutivo externo
homólogo
inferência filogenética
leitura de pontas pareadas

matriz de leitura aberta (ORF)
microarranjo
montagem de sequência
ortólogo
parálogo
parcimônia
pirossequenciamento
projeto genoma
proteoma
pseudogene
pseudogene processado
RNA de interferência (RNAi)
sequência consenso
sequência contig
sequenciamento de RNA
sintenia
supercontig
teste di-híbrido
variação no número de cópias (CNV)
vetor

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Você deseja estudar o desenvolvimento do sistema olfatório (de recepção de cheiro) em camundongo. Você sabe que as células que detectam odores químicos específicos (odorantes) estão localizadas no revestimento das passagens nasais de camundongo. Descreva algumas abordagens para a utilização da genética reversa para estudar a olfação.

Solução

Existem muitas abordagens que podem ser imaginadas. Em relação à genética reversa, você deseja identificar genes candidatos que são expressos no

revestimento das passagens nasais. Em virtude das técnicas da genômica funcional, essa identificação pode ser conquistada por meio da purificação do RNA a partir de células de revestimento da passagem nasal isoladas e da utilização desse RNA como uma sonda de *chips* de DNA que contém sequências que correspondem a todos os mRNA conhecidos em camundongo. Por exemplo, você pode optar por primeiramente examinar os mRNA que estão expressos no revestimento da passagem nasal, mas em nenhum outro local em camundongo, como candidatos importantes para um papel específico na olfação. (Muitas das moléculas importantes também podem apresentar outros papéis em outros locais no corpo, mas você precisa começar em algum ponto.) Alternativamente, você pode optar por iniciar com aqueles genes cujos produtos proteicos são proteínas candidatas à ligação dos próprios odorantes. Independentemente da sua escolha, a próxima etapa seria construir um nocaute direcionado do gene que codifica cada mRNA ou proteína de interesse ou utilizar RNA de interferência para tentar obter fenocópia do fenótipo de perda de função de cada um dos genes candidatos.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Com base na Figura 14.2, por que os fragmentos de DNA sequenciados devem estar sobrepostos para obter uma sequência genômica?
2. O preenchimento dos intervalos no rascunho de sequências genômicas é um desafio importante. Com base na Figura 14.6, as leituras de pontas pareadas de uma biblioteca de fragmentos de 2 kb podem preencher um intervalo de 10 kb?
3. Na Figura 14.9, como as posições dos códons são determinadas?
4. Na Figura 14.10, etiquetas de sequências expressas (EST) estão alinhadas com a sequência genômica. Como as EST são úteis na anotação do genoma?
5. Na Figura 14.10, sequências de cDNA estão alinhadas com a sequência genômica. Como as sequências de cDNA são úteis na anotação do genoma? Os cDNA são mais importantes para as anotações do genoma bacteriano ou

eucariótico?

6. Com base na Figura 14.14 e nas características de elementos ultraconservados, o que você prevê que observaria se injetasse o construto de um gene repórter de rato, ortólogo do elemento ultraconservado *ISL1* em ovócitos fertilizados de camundongo e examinasse a expressão do gene repórter no embrião em desenvolvimento?
7. A Figura 14.17 demonstra regiões sintênicas do cromossomo 11 de camundongo e do cromossomo 17 humano. O que essas regiões sintênicas revelam a respeito do genoma do último ancestral comum de camundongos e humanos?
8. Na Figura 14.18, qual etapa-chave possibilita o sequenciamento do exoma e o distingue do sequenciamento do genoma inteiro?
9. Os genomas de duas linhagens de *E. coli* são comparados na Figura 14.19. Você esperaria que uma terceira linhagem contivesse mais das regiões em azul, em marrom ou vermelho demonstradas na Figura 14.19? Explique.
10. A Figura 14.21 ilustra o sistema di-híbrido baseado em Gal4. Por que as proteínas "isca" fundidas à proteína de ligação ao DNA Gal4 ativam a expressão do gene repórter?

PROBLEMAS BÁSICOS

11. Explique a abordagem que você aplicaria para sequenciar o genoma de uma espécie bacteriana recentemente descoberta.
12. Leituras de sequenciamento terminal de insertos de clones são uma parte rotineira do sequenciamento do genoma. Como a parte central do inserto do clone é obtida?
13. Qual é a diferença entre um contig e um arcabouço?
14. Suspeita-se que dois contigs em particular sejam adjacentes, possivelmente separados por DNA repetitivo. Em uma tentativa de ligá-los, são utilizadas sequências das extremidades como *primers* para tentar preencher o intervalo. Essa abordagem é razoável? Em qual situação ela não atuará?

15. Um segmento de DNA clonado que contém um gene codificador de proteína é marcado radioativamente e utilizado em uma hibridização *in situ* com cromossomos. Foi observada radioatividade ao longo de cinco regiões em diferentes cromossomos. Como esse resultado é possível?

16. Em um experimento de hibridização *in situ*, um determinado clone se ligou apenas ao cromossomo X em um menino sem sintomas de doença. Entretanto, em um menino com distrofia muscular de Duchenne (doença recessiva ligada ao X), ele se ligou ao cromossomo X e a um autossomo. Explique. Esse clone pode ser útil no isolamento do gene da distrofia muscular de Duchenne?

17. Em uma análise genômica na busca por um gene de uma doença específica, observou-se que um gene candidato apresenta uma única substituição de par de bases, que resulta em uma alteração de aminoácido não sinônimo. O que você precisa verificar antes de concluir que identificou o gene causador da doença?

18. Um operador em bactérias é um sítio de ligação?

19. Um determinado cDNA de 2 kb de tamanho hibridizou com oito fragmentos genômicos de 30 kb de tamanho total e que continham duas EST curtas. As EST também foram observadas em dois dos fragmentos genômicos, cada um de 2 kb de tamanho. Esboce uma possível explicação para esses resultados.

20. Um fragmento sequenciado de DNA de *Drosophila* foi utilizado em uma pesquisa BLAST. A melhor (mais próxima) correspondência foi com um gene de quinase de *Neurospora*. Essa correspondência significa que a sequência da *Drosophila* contém um gene de quinase?

21. Em um teste di-híbrido, um determinado gene A proporcionou resultados positivos com dois clones, M e N. Quando M foi utilizado, ele deu positivo com três clones, A, S e Q. O clone N proporcionou apenas um positivo (com A). Desenvolva uma interpretação experimental desses resultados.

22. Você possui as seguintes leituras de sequências de um clone do genoma de

Drosophila melanogaster:
Leitura 1: TGGCCGTGATGGGCAGTTCCGGTG
Leitura 2: TTCCGGTGCCGGAAAGA
Leitura 3: CTATCCGGGCGAACTTTTGGCCG
Leitura 4: CGTGATGGGCAGTTCCGGTG
Leitura 5: TTGGCCGTGATGGGCAGTT
Leitura 6: CGAACTTTTGGCCGTGATGGGCAGTTCC
Utilize essas seis leituras de sequências para criar uma sequência contig dessa parte do genoma de *D. melanogaster.*

23. Algumas vezes, os cDNA tornam-se "monstros"; ou seja, fusões de cópias de DNA de dois mRNA diferentes acidentalmente inseridos de modo adjacente entre si no mesmo clone. Você suspeita de que um clone de cDNA do nematódeo *Caenorhabditis elegans* é um monstro, tendo em vista que a sequência do inserto de cDNA prevê uma proteína com dois domínios estruturais normalmente não observados na mesma proteína. Como você utilizaria a disponibilidade da sequência genômica inteira para avaliar se esse clone de cDNA é um monstro ou não?

24. Ao verificar a sequência do genoma humano, você identifica um gene que apresenta uma região codificadora aparentemente longa, mas existe uma deleção de dois pares de bases que rompe a matriz de leitura.

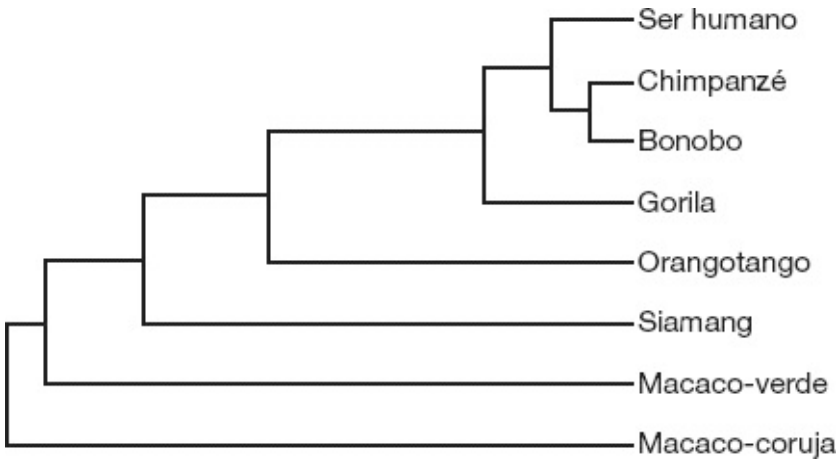
a. Como você determinaria se a deleção estava correta ou foi um erro no sequenciamento?

b. Você observa que existe exatamente a mesma deleção no homólogo do gene de chimpanzé, mas que a matriz de leitura do gene de gorila está intacta. Tendo em vista a filogenia dos grandes primatas, o que você pode concluir a respeito de quando na evolução dos primatas a mutação ocorreu?

25. Ao verificar o genoma do chimpanzé, você observa que ele apresenta três homólogos de um gene em particular, enquanto os seres humanos apresentam apenas dois.

a. Quais são as duas explicações possíveis para essa observação?

b. Como você poderia distinguir entre essas duas possibilidades?



26. O ornitorrinco é um dos poucos mamíferos venenosos. O ornitorrinco-macho apresenta um esporão na pata traseira por meio da qual consegue liberar uma mistura de proteínas venenosas. Observando a filogenia na Figura 14.15, como você determinaria se essas proteínas venenosas são encontradas apenas no ornitorrinco?

27. Você sequenciou o genoma da bactéria *Salmonella typhimurium* e está utilizando uma análise BLAST para identificar semelhanças no genoma de *S. typhimurium* com proteínas conhecidas. Você encontra uma proteína que é 100% idêntica na bactéria *Escherichia coli*. Quando você compara as sequências de nucleotídios dos genes da *S. typhimurium* e de *E. coli*, você observa que suas sequências de nucleotídios são apenas 87% idênticas.

a. Explique essa observação.

b. O que essas observações lhe informam a respeito do mérito das pesquisas de similaridade de nucleotídios *versus* proteínas na identificação de genes correlatos?

28. Para inativar um gene por meio de RNAi, de que informação você necessita? Você precisa da posição do gene-alvo no mapa?

29. Qual é a finalidade de criar uma fenocópia?

30. Qual é a diferença entre genética direta e reversa?

31. Por que o sequenciamento do exoma falharia em identificar uma mutação causadora de uma doença em uma pessoa afetada?

PROBLEMAS DESAFIADORES

32. Você possui as leituras de sequências a seguir de um clone genômico do genoma de *Homo sapiens*:

Leitura 1: ATGCGATCTGTGAGCCGAGTCTTTA
Leitura 2: AACAAAAATGTTGTTATTTTTATTTCAGATG
Leitura 3: TTCAGATGCGATCTGTGAGCCGAG
Leitura 4: TGTCTGCCATTCTTAAAAACAAAAATGT
Leitura 5: TGTTATTTTTATTTCAGATGCGA
Leitura 6: AACAAAAATGTTGTTATT

a. Utilize essas seis leituras de sequências para criar uma sequência contig dessa parte do genoma de *H. sapiens*.

b. Traduza a sequência contig em todas as matrizes de leitura possíveis.

c. Dirija-se à página BLAST do National Center for Biotechnology Information, ou NCBI (http://www.ncbi.nlm.nih.gov/BLAST/, Apêndice B), e verifique se você consegue identificar o gene do qual essa sequência é parte por meio da utilização de cada um dos quadros de leitura como um questionamento para a comparação entre cada proteína (BLASTp).

33. Algumas regiões mensuráveis de diferentes cromossomos do genoma humano apresentam nucleotídios mais de 99% idênticos entre si. Essas regiões foram ignoradas na produção do rascunho da sequência do genoma humano em virtude do seu alto nível de similaridade. Das técnicas discutidas neste capítulo, qual possibilitaria que pesquisadores do genoma identificassem a existência das referidas regiões duplicadas?

34. Alguns éxons no genoma humano são razoavelmente pequenos (menos de 75 pb de comprimento). A identificação desses “microéxons” é difícil, tendo em vista que essas distâncias são muito curtas para utilizar a identificação de ORF ou tendenciosidade de códons de modo confiável para determinar se as pequenas sequências genômicas são verdadeiramente uma parte de um mRNA e um polipeptídio. Quais técnicas de “busca de genes” podem ser utilizadas para tentar avaliar se uma determinada região de 75 pb constitui um éxon?

35. Você está estudando proteínas que apresentam papéis na tradução em camundongo. Por meio da análise BLAST das proteínas previstas do genoma de camundongo, você identifica um conjunto de genes de camundongo que codifica proteínas com sequências semelhantes àquelas de fatores de iniciação de tradução eucariótica conhecidos. Você está interessado em determinar os fenótipos associados às mutações de perda de função desses genes.

a. Você utilizaria abordagens de genética direta ou reversa para identificar essas mutações?

b. Resuma brevemente duas abordagens diferentes que você pode utilizar para procurar fenótipos de perda de função em um desses genes.

36. Todo o genoma da levedura *Saccharomyces cerevisiae* foi sequenciado. Esse sequenciamento levou à identificação de todas as matrizes de leitura abertas (ORF, sequências do tamanho do gene com sinais de início e término da tradução apropriados) no genoma. Alguns desses ORF são genes previamente conhecidos com funções estabelecidas; entretanto, o restante é de matrizes de leitura incógnitas (URF). Para deduzir as possíveis funções dos URF, eles estão sendo convertidos sistematicamente, um por vez, em alelos nulos por meio de técnicas de nocaute *in vitro*. Os resultados são como segue:

15% são letais quando nocauteados.

25% demonstram algum fenótipo mutante (morfologia alterada, nutrição alterada e assim por diante).

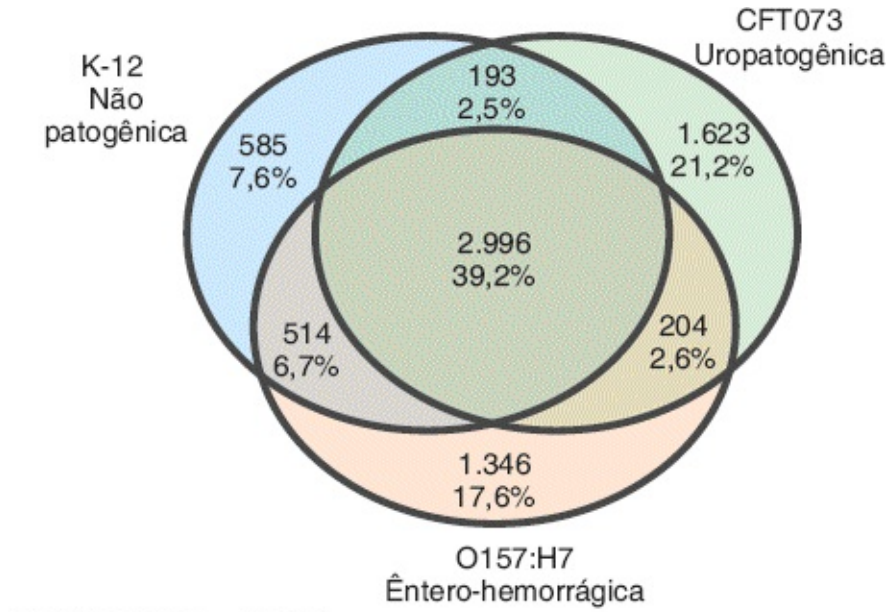
60% absolutamente não demonstram fenótipo mutante detectável e se assemelham ao tipo selvagem.

Explique a possível base genética molecular dessas três categorias de mutantes, inventando exemplos quando possível.

37. Diferentes linhagens de *E. coli* são responsáveis por infecções êntero-hemorrágicas e urinárias. Com base nas diferenças entre a linhagem K-12 benigna e a linhagem O157:H7 êntero-hemorrágica, você preveria que existem diferenças genômicas óbvias:

a. Entre K-12 e as linhagens uropatogênicas?

b. Entre O157:H7 e as linhagens uropatogênicas?



Proteínas totais = 7.638
2.996 (39,2%) em todas as 3
911 (11,9%) em 2 de 3
3.554 (46,5%) em 1 de 3

c. O que poderia explicar as diferenças par a par observadas no conteúdo do genoma?

d. Como a função de genes específicos de linhagens pode ser testada?

[1]M. Johnson e K. Gallagher, “A Baffling Illness”, *Milwaukee Journal Sentinel*. Publicado em 10 de dezembro de 2010. Acesso em 5 de março de 2014.

CAPÍTULO 15

Genoma Dinâmico | Elementos de Transposição

Grãos em uma espiga de milho. Os grãos manchados nesta espiga resultam da interação de um elemento genético móvel (um elemento de transposição) com um gene de milho cujo produto é necessário para a pigmentação. (Cliff Weil e Susan Wessler.)

TÓPICOS

15.1 Descoberta dos elementos de transposição no milho
15.2 Elementos de transposição em procariotos
15.3 Elementos de transposição em eucariotos
15.4 O genoma dinâmico | Mais elementos de transposição do que jamais se imaginou

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Descrever como os elementos de transposição foram primeiro descobertos geneticamente no milho e em seguida foram primeiro isolados molecularmente de *E. coli*
- Descrever como os elementos de transposição participam na dispersão de bactérias resistentes a antibióticos
- Comparar e contrastar os dois mecanismos principais utilizados pelos elementos para a transposição
- Fornecer motivos para explicar como a espécie humana sobrevive, sendo mais de 50% de nosso genoma derivados de elementos de transposição
- Descrever os mecanismos utilizados por genomas hospedeiros para reprimir a difusão de alguns elementos de transposição
- Descrever as estratégias utilizadas pelos elementos de transposição para evitar os mecanismos de repressão dos hospedeiros.

Um menino nasce com uma doença que torna o seu sistema imune ineficiente. Os testes diagnósticos determinam que ele apresenta um distúrbio genético recessivo denominado SCID (doença por imunodeficiência combinada grave), mais comumente conhecida como *doença do menino da bolha*. Essa doença é causada por uma mutação no gene que codifica a enzima sanguínea adenosina desaminase (ADA). Como resultado da perda dessa enzima, as células precursoras que originam um dos tipos celulares do sistema imune estão ausentes. Tendo em vista que esse menino não apresenta a capacidade de combater infecções, ele precisa viver em um ambiente completamente isolado e estéril — ou seja, uma bolha na qual o ar é filtrado para tornar-se estéril (Figura 15.1). Nenhuma terapia medicamentosa ou outra terapia convencional está disponível para tratar essa doença. A realização, no menino, de um transplante de

tecido contendo as células precursoras de outra pessoa não funcionaria na maioria dos casos, tendo em vista que uma compatibilidade tecidual precisa entre o doador e o paciente é extremamente rara. Consequentemente, as células doadoras acabariam por criar uma resposta imune contra os tecidos do próprio menino (doença enxerto *versus* hospedeiro).

Nas duas últimas décadas, foram desenvolvidas técnicas que oferecem a possibilidade de um diferente tipo de terapia com transplante — a **terapia gênica** — que poderia ajudar pessoas com SCID e outras doenças genéticas incuráveis. Em relação à SCID, um gene ADA normal é "transplantado" para células do sistema imunológico de um paciente, possibilitando, assim, que essas células sobrevivam e funcionem normalmente. Nos estudos iniciais de terapia gênica humana, os cientistas modificaram um tipo de vírus denominado retrovírus em laboratório, de modo que ele pudesse inserir a si próprio em um gene ADA normal nos cromossomos das células imunes coletadas de pacientes com SCID. Neste capítulo, você verá que os retrovírus apresentam muitas propriedades biológicas em comum com um tipo de elemento móvel denominado *retrotranspóson*, que está presente em nosso genoma e nos genomas da maior parte dos eucariotos. As lições aprendidas a respeito do comportamento dos retrotranspósons e de outros elementos móveis a partir de organismos-modelo, como as leveduras, são fontes de valiosas descobertas de uma nova geração de vetores biológicos para a terapia gênica humana.

FIGURA 15.1 Um paciente com SCID deve viver em uma bolha protetora. (© Bettmann/Corbis.)

Com início na década de 1930, estudos genéticos no milho produziram resultados que superaram muito o quadro da genética clássica dos genes localizados apenas em *loci* fixos nos cromossomos. A literatura científica começou a transmitir relatos que sugeriam que determinados elementos genéticos presentes nos cromossomos de algum modo poderiam se mover de um local para outro. Esses achados foram considerados com ceticismo durante muitos anos, mas atualmente está claro que tais elementos móveis são muito disseminados na natureza.

Vários termos curiosos (alguns dos quais ajudam a descrever suas respectivas propriedades) foram aplicados a esses elementos genéticos: elementos controladores, genes saltadores, genes móveis, elementos móveis e transpósons. Aqui, utilizamos os termos *elementos transponíveis* ou *de transposição* e *elementos móveis*, que abrangem toda a família de tipos. Os elementos de transposição conseguem mover-se até novas posições dentro do mesmo cromossomo ou até mesmo para um cromossomo diferente. Eles foram detectados geneticamente em organismos-modelo, tais como *E. coli*, milho, levedura, *C.*

elegans e *Drosophila*, por meio de mutações que eles produzem quando se inserem em genes e os inativam.

O sequenciamento do DNA de genomas de uma diversidade de microrganismos, plantas e animais indica que os elementos de transposição existem em praticamente todos os organismos. Surpreendentemente, eles são, de longe, o maior componente do genoma humano, responsáveis por quase 50% de nossos cromossomos. Apesar da sua abundância, o papel genético normal desses elementos não é conhecido com certeza.

Em seus estudos, os cientistas são capazes de explorar a capacidade dos elementos de transposição de se inserirem em novos sítios no genoma. Os elementos de transposição modificados em tubo de ensaio são ferramentas valiosas, tanto em procariotos quanto em eucariotos, para o mapeamento genético, criação de mutantes, clonagem de genes e mesmo para a produção de organismos transgênicos. Reconstruiremos algumas das etapas na evolução de nossa atual compreensão sobre os elementos de transposição. Ao fazer isso, revelaremos os princípios que orientam essas fascinantes unidades genéticas.

15.1 Descoberta dos elementos de transposição no milho

Experimentos de McClintock | O elemento *Ds*

Na década de 1940, Barbara McClintock fez uma descoberta incrível enquanto estudava os grãos coloridos do chamado milho indiano (ver Organismo-modelo, Milho, adiante). O milho tem 10 cromossomos, numerados do maior (1) até o menor (10). Enquanto analisava a quebra dos cromossomos desse milho, McClintock observou alguns fenômenos incomuns. Ela observou que, em uma linhagem de milho, o cromossomo 9 quebrava muito frequentemente em um local particular, ou *locus* (Figura 15.2). Ela determinou que a quebra do cromossomo nesse *locus* ocorria em virtude da presença de dois fatores genéticos: um fator, que ela denominou ***Ds*** (em referência à ***Dissociação***), e estava localizado no lugar da quebra, e outro fator, não ligado, necessário para "ativar" a quebra do cromossomo 9 no *locus Ds*. McClintock denominou esse segundo fator ***Ac*** (em

referência a ***Ativador***).

McClintock começou a suspeitar de que *Ac* e *Ds* fossem de fato elementos genéticos móveis quando ela descobriu ser impossível mapear *Ac*. Em algumas plantas, ele mapeava em uma posição; em outras plantas da mesma linhagem, ele mapeava em posições diferentes. Como se esse mapeamento variável não fosse uma curiosidade suficiente, grãos raros, com fenótipos marcadamente diferentes, podiam ser derivados da linhagem original que sofreu quebras frequentes no cromossomo 9. Um desses fenótipos era um grão incolor raro que continha manchas pigmentadas.

A Figura 15.3 compara o fenótipo da linhagem com a quebra cromossômica com o fenótipo de uma dessas linhagens derivadas. Em relação à linhagem com quebra cromossômica, um cromossomo que se quebra em *Ds* ou perto dele perde a sua extremidade que contém alelos do tipo selvagem dos genes *C, Sh* e *Wx*. No exemplo demonstrado na Figura 15.3 A, ocorreu uma quebra em uma única célula, que se dividiu mitoticamente até produzir o grande setor de tecido mutante (*c sh wx*). A quebra pode ocorrer muitas vezes em um único grão, mas cada setor de tecido demonstrará a perda de expressão de todos os três genes. Contrariamente, cada novo derivado afetava a expressão de apenas um único gene. Um derivado que afetava a expressão apenas do gene *C* de pigmento está demonstrado na Figura 15.3 B. Nesse exemplo, manchas pigmentadas apareceram em um fundo de grão incolor. Embora a expressão de *C* tenha sido alterada desse modo estranho, a expressão de *Sh* e *Wx* era normal e o cromossomo 9 deixou de apresentar quebras frequentes.

Para explicar os novos derivados, McClintock formulou a hipótese de que *Ds* havia se movido de um sítio perto do centrômero para o gene *C* localizado próximo da extremidade telomérica. Nessa nova localização, *Ds* evita a expressão de *C*. A inativação do gene *C* explica as partes incolores do grão, mas o que explica o aparecimento das manchas pigmentadas? O grão manchado é um exemplo de um **fenótipo instável**. McClintock concluiu que tais fenótipos instáveis resultavam da movimentação, ou transposição, de *Ds* para longe do gene *C*. Ou seja, o grão inicia o desenvolvimento com um gene *C* que sofreu mutação pela inserção de *Ds*. Entretanto, em algumas células do grão, *Ds* deixa o gene *C*,

possibilitando que o fenótipo mutante seja revertido para o tipo selvagem e produza o pigmento na célula original e em todos os seus descendentes mitóticos. Existem grandes manchas coloridas quando *Ds* deixa o gene *C* inicialmente no desenvolvimento do grão (tendo em vista que existem mais descendentes mitóticos), enquanto existem manchas pequenas quando *Ds* deixa o gene *C* posteriormente no desenvolvimento do grão. Fenótipos mutantes instáveis que revertem ao tipo selvagem são uma indicação da participação de elementos móveis.

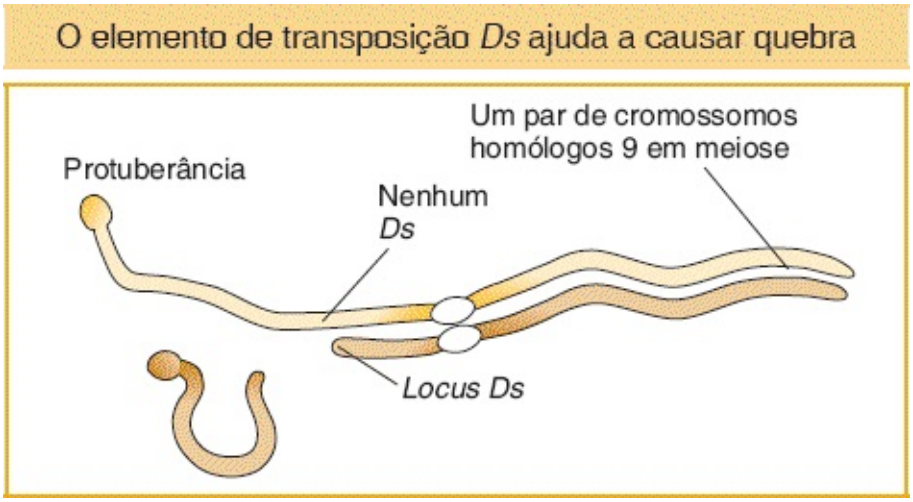


FIGURA 15.2 O cromossomo 9 de milho quebra no *locus DS*, em que o elemento de transposição *Ds* foi inserido.

Elementos autônomos e não autônomos

Qual é a relação entre *Ac* e *Ds*? Como eles interagem com os genes e os cromossomos para produzir esses fenótipos interessantes e incomuns? Essas questões foram respondidas por meio de análise genética adicional. As interações de *Ds*, *Ac* e o gene *C* de pigmento são utilizadas como um exemplo na Figura 15.4. Aqui, *Ds* é demonstrado como um segmento de DNA que tem o gene *C* inativado pela inserção em sua região codificadora. O alelo que carreia o inserto é denominado *c-mutável(Ds)* ou *c-m(Ds)* abreviadamente. Uma linhagem com *c-m(Ds)* e nenhum *Ac* apresenta grãos incolores, tendo em vista que *Ds* não consegue se mover; ele está preso no gene *C*. Uma linhagem com *c-m(Ds)* e *Ac*

apresenta grãos manchados, tendo em vista que *Ac* ativa *Ds* em algumas células para deixar o gene *C*, restaurando, assim, a função do gene. Diz-se que o elemento que sai é **excisado** do cromossomo, ou **transposto**.

Foram isoladas outras linhagens nas quais o próprio elemento *Ac* havia se inserido no gene *C* (denominado *c-m(Ac)*). Contrariamente ao alelo *c-m(Ds)*, que é instável apenas quando *Ac* está no genoma, *c-m(Ac)* é sempre instável. Além disso, McClintock observou que, em raras ocasiões, um alelo do tipo *Ac* poderia ser transformado em um alelo do tipo *Ds*. Essa transformação ocorria em virtude da geração espontânea de um elemento *Ds* a partir do elemento *Ac* inserido. Em outras palavras, *Ds* é, com todas as probabilidades, uma versão mutada e incompleta do próprio *Ac*.

Diversos sistemas como *Ac/Ds* foram encontrados por McClintock e outros geneticistas que trabalhavam com milho. Dois outros sistemas são o *Dotted* ([*Dt*], descoberto por Marcus Rhoades) e o *Supressor/mutator* ([*Spm*], descoberto independentemente por McClintock e Peter Peterson, que o denominaram *Enhancer/Inhibitor* [*En/In*]). Além disso, conforme você verá nas seções a seguir, foram isolados elementos com comportamento genético semelhante em bactérias, plantas e animais.

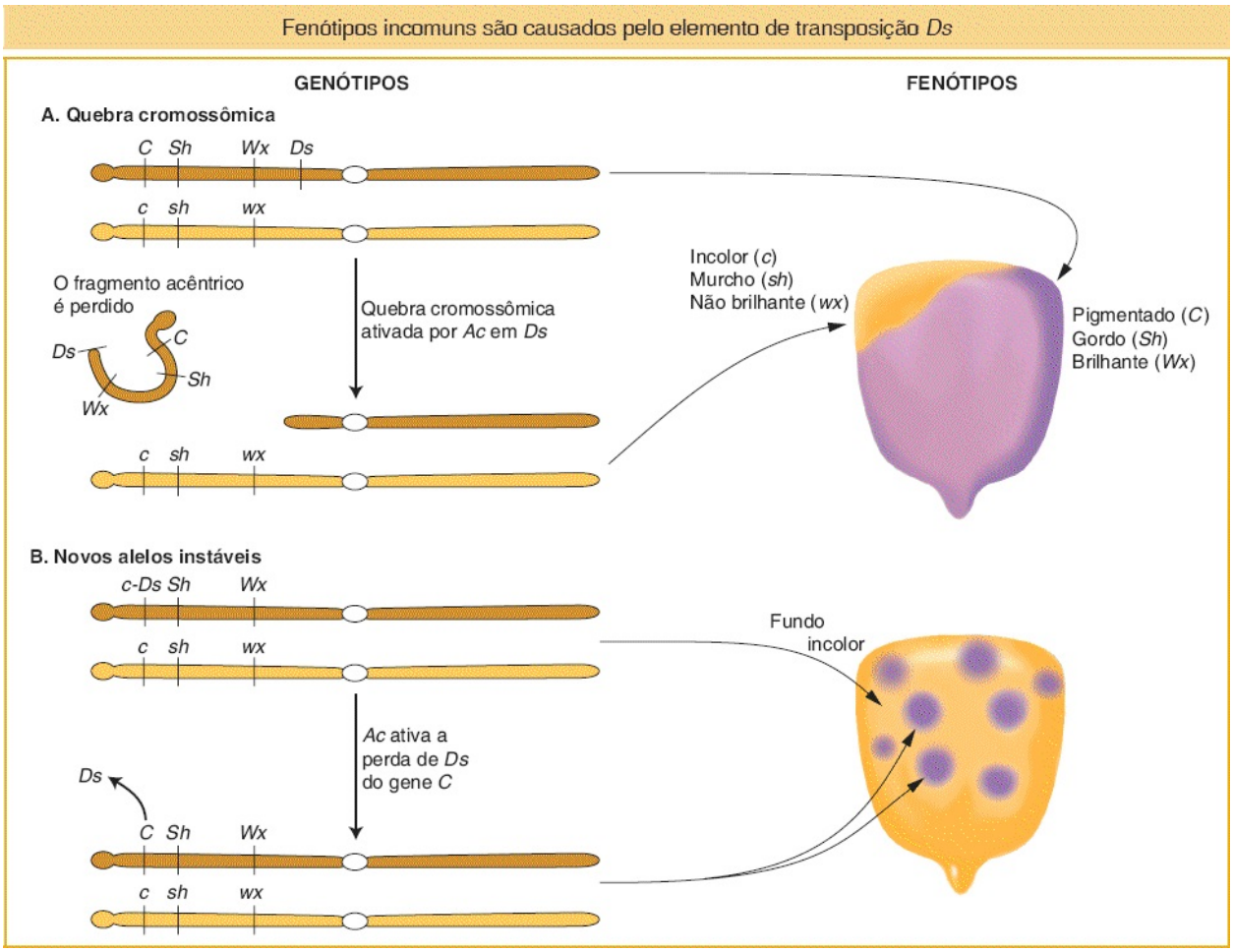


FIGURA 15.3 Novos fenótipos no milho são produzidos por meio da movimentação do elemento de transposição *Ds* no cromossomo 9. **A.** Um fragmento cromossômico é perdido por meio da quebra no *locus* de *Ds*. Os alelos recessivos no cromossomo homólogo são expressos, produzindo o setor incolor no grão. **B.** A inserção de *Ds* no gene *C* (*parte superior*) cria células de grão de milho incolores. A excisão de *Ds* do gene *C* por meio da ação de *Ac* nas células e nos seus descendentes mitóticos possibilita que a cor seja novamente expressa, produzindo o fenótipo manchado.

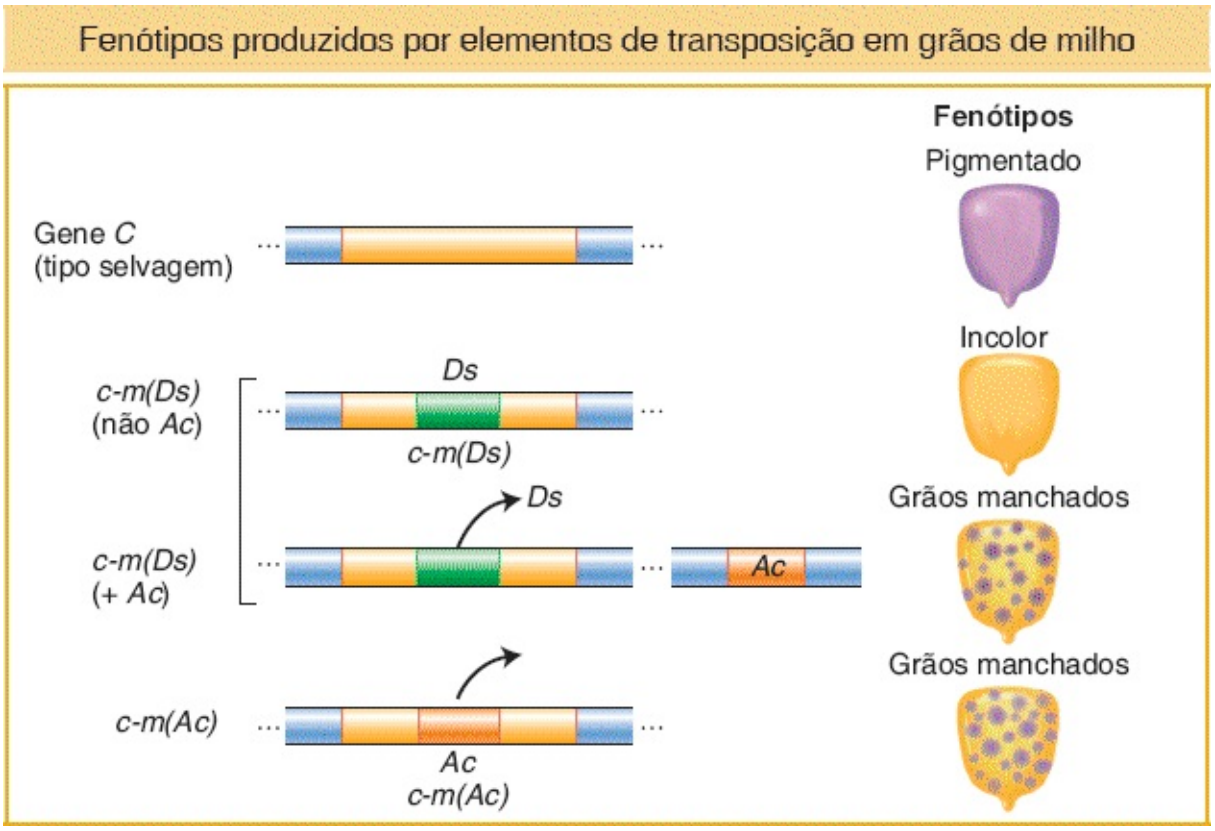


FIGURA 15.4 As manchas do grão são controladas pela inserção e pela excisão de elementos *Ds* ou *Ac* no gene *C* que controla o pigmento

ORGANISMO-MODELO *Milho*

O milho, *Zea mays*, é membro da família das gramíneas. As gramíneas — que incluem também o arroz, o trigo e a cevada — são a fonte mais importante de calorias para a humanidade. O milho foi domesticado a partir da grama selvagem teosinto pelos indígenas americanos no México e na América Central, e foi introduzido na Europa por Colombo em seu retorno do Novo Mundo.

Na década de 1920, Rollins A. Emerson estabeleceu um laboratório na Cornell University para estudar a genética de traços do milho, incluindo a cor dos grãos, ideais para a análise genética. Além disso, a separação física das flores masculinas e femininas na borla e na espiga, respectivamente, tornava os cruzamentos genéticos controlados fáceis de ser realizados. Entre os excepcionais geneticistas atraídos para o laboratório de Emerson estavam

Marcus Rhoades, Barbara McClintock e George Beadle (ver Capítulo 6). Antes do advento da biologia molecular e do surgimento de microrganismos como organismos-modelo, os geneticistas realizavam as análises microscópicas dos cromossomos e relacionavam o seu comportamento à segregação dos traços. Os grandes cromossomos da ponta da raiz do milho e da glândula salivar de *Drosophila* os tornaram os organismos de escolha para as análises citogenéticas. Os resultados desses estudos iniciais levaram a uma compreensão sobre o comportamento dos cromossomos durante a meiose e a mitose, incluindo eventos tais como a recombinação e as consequências de quebras cromossômicas, tais como inversões, translocações e duplicação.

O laboratório de milho de Rollins A. Emerson na Cornell University, 1929. Em pé, da esquerda para a direita: Charles Burnham, Marcus Rhoades, Rollins Emerson e Barbara McClintock. Agachado está George Beadle. Tanto McClintock quanto Beadle receberam um Prêmio Nobel. (Department of Plant Breeding, Cornell University.)

Análise dos cromossomos de milho, naquela época e atualmente. Os cromossomos de milho são grandes e facilmente visualizados por microscopia óptica. Uma imagem de Marcus Rhoades (1952). (James A. Birchler, R. Kelly Dawe e John F. Doebley, "Perspectives Anecdotal, Historical and Critical Commentaries on Genetics: Marcus Rhoades, Preferential Segregation and Meiotic Drive." © 2003 Genetics Society of America, p. 836.)

O milho ainda atua como um organismo-modelo genético. Biólogos moleculares continuam a explorar seus belos cromossomos do paquíteno com novas sondas de anticorpos, e utilizaram a sua riqueza de elementos de transposição geneticamente bem-caracterizados como ferramentas para identificar e isolar genes importantes.

O comportamento genético comum desses elementos levou os geneticistas a propor novas categorias para todos os elementos. *Ac* e elementos com propriedades genéticas semelhantes atualmente são denominados **elementos autônomos**, tendo em vista que não necessitam de outros elementos para a sua mobilidade. De modo semelhante, *Ds* e elementos com propriedades genéticas semelhantes são denominados **elementos não autônomos**. Uma *família* de

elementos é composta por um ou mais elementos autônomos e pelos membros não autônomos que podem ser mobilizados. Os elementos autônomos codificam a informação necessária para a sua própria movimentação e para a movimentação de elementos não autônomos não ligados no genoma. Tendo em vista que os elementos não autônomos não codificam as funções necessárias para sua própria movimentação, eles não conseguem se movimentar, exceto se um elemento autônomo de sua família estiver presente em alguma outra parte do genoma.

A Figura 15.5 demonstra um exemplo dos efeitos dos transpósons em uma rosa.

> CONCEITO-CHAVE Os elementos de transposição em milho podem inativar um gene no qual eles estão localizados, causar quebras cromossômicas e se transpor para novos locais dentro do genoma. Os elementos autônomos podem realizar essas funções sem auxílio; os elementos não autônomos só podem se transpor apenas com o auxílio de um elemento autônomo em alguma outra parte do genoma.

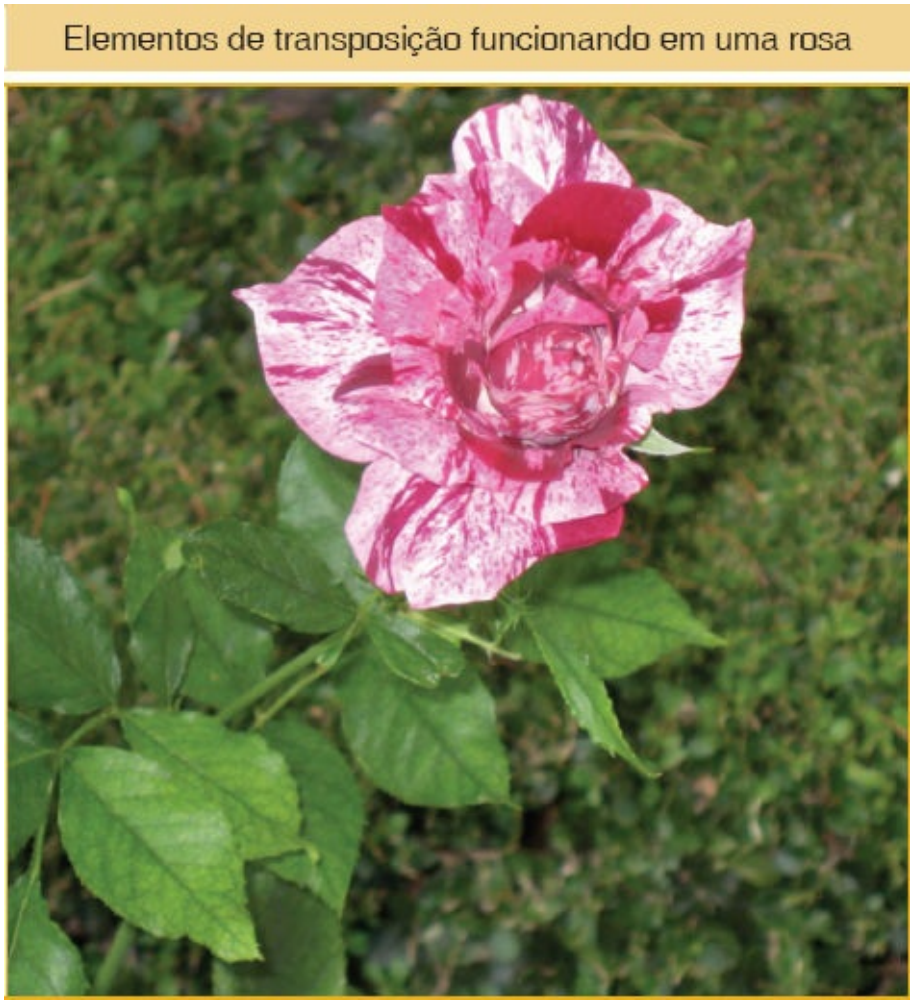


FIGURA 15.5 O mosaicismo é causado pela excisão de elementos de transposição em rosas. A inserção de um elemento de transposição rompe a produção de pigmento, resultando em flores brancas. A excisão do elemento de transposição restaura a produção de pigmento, resultando em setores de tecido floral vermelho. (Susan Wessler.)

Elementos de transposição | Apenas no milho?

Embora os geneticistas tenham aceitado a descoberta de McClintock dos elementos de transposição em milho, muitos relutaram em considerar a possibilidade de que elementos semelhantes estivessem presentes nos genomas de outros organismos. Sua existência em todos os organismos implicaria o fato de os genomas serem inerentemente instáveis e dinâmicos. Essa visão era inconsistente com o fato de que os mapas genéticos de membros da mesma espécie eram iguais. Afinal, se genes podem ser geneticamente mapeados em uma localização

cromossômica precisa, esse mapeamento não indica que eles não estão se movimentando?

Tendo em vista que McClintock era uma geneticista altamente respeitada, seus resultados não foram questionados. Em vez disso, sua relevância em relação a outros organismos foi questionada por outras pessoas, que argumentavam que o milho não é um organismo natural: ele é uma planta de cultivo que é produto da seleção humana e da domesticação. Essa visão foi mantida até a década de 1960, quando os primeiros elementos de transposição foram isolados do genoma de *E. coli* e estudados no nível da sequência de DNA. Os elementos de transposição foram subsequentemente isolados dos genomas de muitos organismos, incluindo *Drosophila* e levedura. Quando se tornou aparente que os elementos de transposição são um componente significativo dos genomas da maior parte e talvez de todos os organismos, Barbara McClintock foi reconhecida por sua importante descoberta e recebeu o Prêmio Nobel de Fisiologia ou Medicina em 1983.

15.2 Elementos de transposição em procariotos

A descoberta genética dos elementos de transposição levou a muitas questões a respeito da semelhança na sequência de DNA de tais elementos e como eles são capazes de se movimentar de um local para outro no genoma. Todos os organismos os apresentam? Todos os elementos se assemelham ou existem classes diferentes de elementos de transposição? Se existem muitas classes de elementos, eles podem coexistir em um genoma? O número de elementos de transposição no genoma varia de espécie para espécie? A natureza molecular dos elementos genéticos de transposição foi compreendida pela primeira vez em bactérias. Portanto, continuaremos essa história ao examinar os estudos originais realizados com procariotos.

Existem dois tipos amplos de elementos de transposição em bactérias:

- Sequências curtas denominadas *elementos IS*, que conseguem se mover até novas posições, mas que não carreiam outros genes além daqueles necessários

para seu movimento
- Sequências mais longas denominadas *transpósons*, que não apenas carreiam os genes de que necessitam para seu movimento, mas que também carreiam outros genes.

Sequências de inserção bacteriana

As sequências de inserção, ou **elementos de sequência de inserção (IS)**, são segmentos de DNA bacteriano que conseguem se mover de uma posição em um cromossomo para uma posição diferente no mesmo cromossomo ou em um cromossomo diferente. Quando os elementos IS aparecem na parte intermediária dos genes, eles interrompem a sequência codificadora e inativam a expressão daquele gene. Em virtude do seu tamanho e, em alguns casos, da presença de sinais de término de transcrição e tradução no elemento IS, os elementos IS também conseguem bloquear a expressão de outros genes no mesmo óperon, se aqueles genes estiverem *downstream* do sítio de inserção. Os elementos IS foram observados pela primeira vez em *E. coli* no óperon *gal* — um grupo de três genes que participam do metabolismo do açúcar galactose.

Identificação de elementos IS bem-definidos. Observou-se que diversos mutantes *gal*$^{-}$ de *E. coli* contêm grandes inserções de DNA no óperon *gal*. Esse achado naturalmente levou à próxima questão: os segmentos de DNA que se inserem nos genes são meramente fragmentos aleatórios de DNA ou são entidades genéticas distintas? A resposta para essa questão teve origem nos resultados de experimentos de hibridização que demonstram que muitas mutações de inserção diferentes são causadas por um pequeno conjunto de sequências de inserção. Esses experimentos são realizados com a utilização de fagos λd*gal* que contêm o óperon *gal*$^{-}$ de diversas linhagens mutantes de *gal* independentemente isoladas. Partículas de fagos individuais são isoladas a partir das linhagens e seu DNA é utilizado para sintetizar RNA radioativo *in vitro*. Observa-se que determinados fragmentos desse RNA hibridizam com o DNA de outras mutações *gal*$^{-}$ que contêm grandes inserções de DNA, mas não com DNA do tipo selvagem. Esses resultados foram interpretados como evidência de que mutantes *gal*

independentemente isolados contêm o mesmo pedaço extra do DNA. Esses fragmentos particulares de RNA também hibridizam com o DNA de outros mutantes que contêm inserções IS em outros genes, demonstrando que o mesmo trecho de DNA pode inserir-se em diferentes locais no cromossomo bacteriano.

Estrutura dos elementos IS. Com base em seus padrões de hibridização cruzada, foi identificado um número de elementos IS distintos. Uma sequência, denominada IS1, é o segmento de 800 pb identificado em *gal*. Outra sequência, denominada IS2, apresenta o comprimento de 1.350 pb. Embora os elementos IS sejam diferentes na sequência de DNA, eles apresentam diversas características em comum. Por exemplo, todos os elementos IS codificam uma proteína, denominada **transposase**, que é uma enzima necessária para a movimentação dos elementos IS de um local no cromossomo para outro. Além disso, todos os elementos IS começam e terminam com sequências curtas de repetições invertidas, que são necessárias para a sua mobilidade. A transposição dos elementos IS e outros elementos genéticos móveis será considerada posteriormente no capítulo.

O genoma do tipo selvagem padrão de *E. coli* é rico em elementos IS: ele contém oito cópias de IS1, cinco cópias de IS2 e cópias de outros tipos de IS não tão bem-estudados. Tendo em vista que os mesmos elementos IS apresentam sequências idênticas, eles são sítios nos quais podem ocorrer *crossovers*. Por exemplo, a recombinação entre o plasmídio fator F e o cromossomo de *E. coli* para formar linhagens *Hfr* é o resultado de um único *crossover* entre o elemento IS1 localizado no plasmídio e um elemento IS1 localizado no cromossomo (ver Figura 5.18). Tendo em vista que existem múltiplos elementos IS1, o fator F pode se inserir em múltiplos sítios.

> CONCEITO-CHAVE O genoma bacteriano contém segmentos de DNA, denominados elementos IS, que conseguem se mover de uma posição no cromossomo até uma posição diferente no mesmo cromossomo ou em um cromossomo diferente.

Transpósons procarióticos

No Capítulo 5, você aprendeu a respeitos dos **fatores R**, plasmídios que carreiam genes que codificam a resistência a diversos antibióticos. Esses fatores R (em referência à resistência), também conhecidos como plasmídios R, são transferidos rapidamente com a conjugação celular, de um modo muito semelhante ao fator F em *E. coli*.

Os fatores R são justamente os primeiros de muitos fatores semelhantes do tipo F a serem descobertos. Observou-se que os fatores R carreiam muitos tipos diferentes de genes em bactérias. Em particular, os fatores R coletam genes que conferem resistência a diferentes antibióticos. Como eles adquirem suas novas capacidades genéticas? Ocorre que os genes de resistência a fármacos estão localizados em um elemento genético móvel denominado **transpóson (Tn)**. Existem dois tipos de transpósons bacterianos. Os **transpósons compostos** contêm uma diversidade de genes que estão localizados entre dois elementos IS quase idênticos, orientados em sentido oposto (Figura 15.6 A) e, como tal, formam o que é denominada uma sequência de **repetição invertida**. A transposase codificada por um dos dois elementos IS é necessária para catalisar o movimento do transpóson inteiro. Um exemplo de um transpóson composto é o Tn10, demonstrado na Figura 15.6 A. O Tn10 carreia um gene que confere a resistência ao antibiótico tetraciclina e é flanqueado por dois elementos IS10 em orientação oposta. Os elementos IS que compõem os transpósons compostos não são capazes de ser transpostos por conta própria em virtude de mutações em suas repetições invertidas.

Os **transpósons simples** também são compostos por genes bacterianos flanqueados por sequências de repetições invertidas, mas essas sequências são curtas (< 50 pb) e não codificam a enzima transposase que é necessária para a transposição. Portanto, a sua mobilidade não ocorre em virtude de uma associação com elementos IS. Em vez disso, os transpósons simples codificam a sua própria transposase na região entre as sequências de repetições invertidas, além de carrear os genes bacterianos. Um exemplo de transpóson simples é o Tn3, demonstrado na Figura 15.6 B.

Para uma revisão, os elementos IS são sequências móveis curtas que codificam apenas aquelas proteínas necessárias para a sua mobilidade. Os transpósons compostos e os transpósons simples contêm genes adicionais que conferem funções novas para as células bacterianas. Sejam compostos ou simples, os transpósons normalmente são denominados apenas transpósons, e transpósons diferentes são designados Tn1, Tn2, Tn505 e assim por diante.

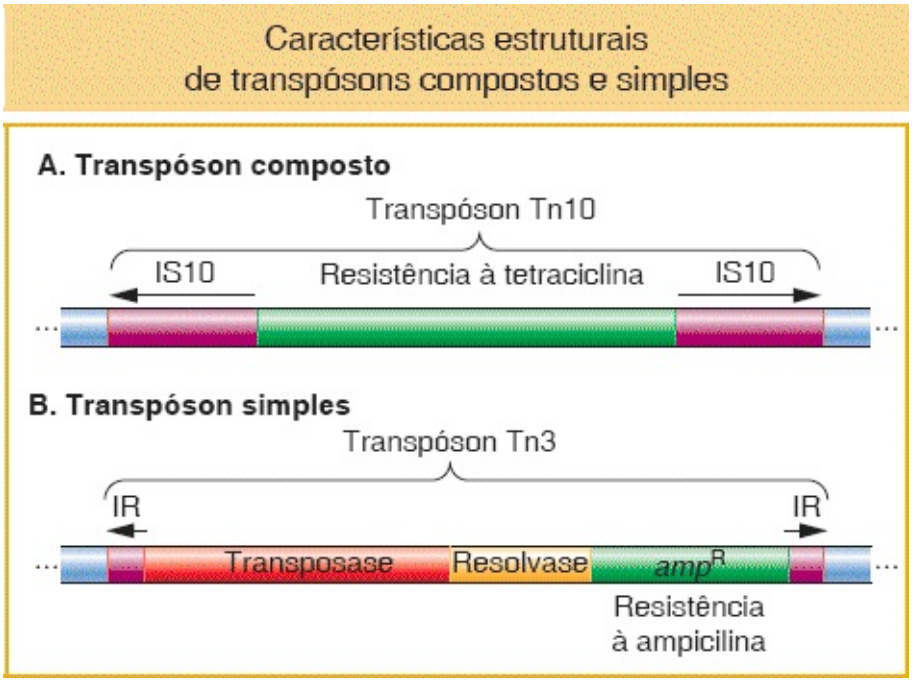


FIGURA 15.6 A. Tn10, um exemplo de um transpóson composto. Os elementos IS estão inseridos em orientação oposta e formam repetições invertidas (IR). Cada elemento IS carreia uma transposase, mas apenas uma normalmente é funcional. **B.** Tn3, um exemplo de um transpóson simples. Repetições invertidas curtas não contêm transposase. Em vez disso, os transpósons simples codificam sua própria transposase. A resolvase é uma proteína que promove recombinação e que resolve os cointegrados (ver Figura 15.9).

Um transpóson pode pular de um plasmídio para um cromossomo bacteriano ou de um plasmídio para outro plasmídio. Desse modo, são gerados plasmídios resistentes a múltiplos fármacos. A Figura 15.7 é um diagrama composto de um fator R, que indica os diversos locais nos quais os transpósons podem estar localizados. Em seguida, consideramos a questão de como ocorrem tais eventos de **transposição** ou mobilização.

> CONCEITO-CHAVE Os transpósons foram originalmente detectados como elementos genéticos móveis que conferem a resistência a fármacos. Muitos desses elementos são compostos por elementos IS que flanqueiam um gene que codifica a resistência a fármacos. Essa organização promove a disseminação de bactérias resistentes a fármacos por meio da facilitação do movimento do gene de resistência do cromossomo de uma bactéria resistente para um plasmídio que pode ser conjugado a outra linhagem bacteriana (suscetível).

Mecanismo de transposição

Conforme já declarado, a movimentação de um elemento de transposição depende da ação de uma transposase. Essa enzima desempenha papéis-chave nos dois estágios da transposição: excisão (saída) do local original e inserção em um novo local.

Excisão do local original. A maior parte dos elementos de transposição em procariotos (e em eucariotos) emprega um de dois mecanismos de transposição, denominados **replicativo** e **conservativo** (não replicativo), conforme ilustrado na Figura 15.8. Na via replicativa (conforme demonstrada em relação ao Tn3), uma nova cópia do elemento de transposição é gerada no evento de transposição. Os resultados da transposição são que uma cópia aparece no novo sítio e uma cópia permanece no sítio antigo. Na via conservativa (conforme demonstrada em relação ao Tn10), não existe replicação. Em vez disso, o elemento é excisado do cromossomo ou do plasmídio e é integrado ao novo sítio. A via conservativa também é denominada "**corte e colagem**".

Transposição replicativa. Tendo em vista que esse mecanismo é um pouco complicado, ele será descrito aqui em mais detalhes. Assim como a Figura 15.8 ilustra, uma cópia de Tn3 é produzida a partir de uma cópia única inicial, produzindo duas cópias de Tn3.

A Figura 15.9 demonstra os detalhes dos intermediários na transposição do Tn3

de um plasmídio (o doador) para outro plasmídio (o alvo). Durante a transposição, os plasmídios doador e receptor são temporariamente fundidos, formando um plasmídio duplo. A formação desse intermediário é catalisada pela transposase codificada por Tn3, que realiza cortes no filamento único nas duas extremidades de Tn3 e cortes desencontrados na sequência-alvo e une as extremidades livres, formando um círculo fundido, denominado **cointegrado**. O elemento de transposição é duplicado no evento de fusão. O cointegrado em seguida é resolvido por meio de um evento semelhante à recombinação, que transforma um cointegrado em dois círculos menores, deixando uma cópia do elemento de transposição em cada plasmídio. O resultado é que uma cópia permanece no local original do elemento, enquanto a outra é integrada em uma nova posição genômica.

Transposição conservativa. Alguns transpósons, tais como o Tn10, são excisados do cromossomo e integrados no DNA-alvo. Nesses casos, o DNA do elemento não é replicado e o elemento é perdido a partir do sítio do cromossomo original (ver Figura 15.8). Assim como a transposição replicativa, essa reação é iniciada pela transposase codificada pelo elemento, que realiza cortes nas extremidades do transpóson. Entretanto, contrariamente à transposição replicativa, a transposase corta o elemento fora do sítio do doador. Em seguida, ela realiza um corte desencontrado em um sítio-alvo e insere nele o elemento. Revisaremos esse mecanismo em mais detalhes em uma discussão sobre a transposição de elementos de transposição eucarióticos, incluindo a família de *Ac/Ds* de milho.

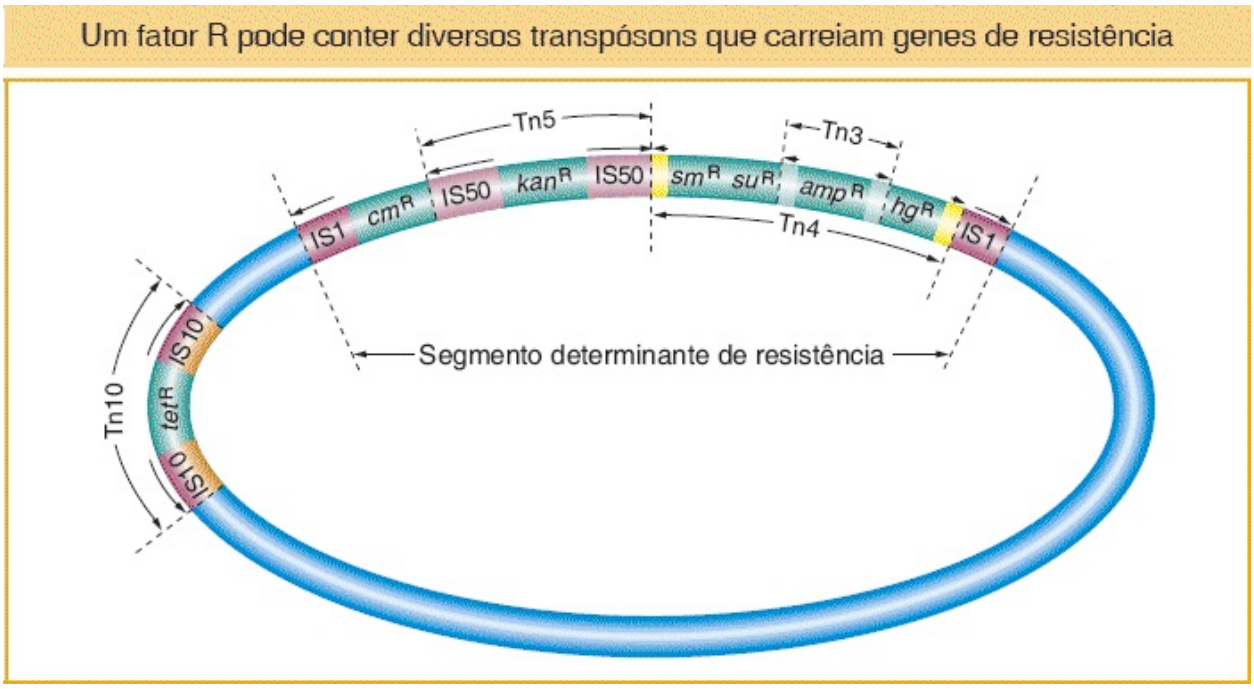


FIGURA 15.7 Mapa esquemático de um plasmídio com diversas inserções de transpósons simples e compostos que carreiam genes de resistência. As sequências dos plasmídios estão em *azul*. Estão demonstrados os genes que codificam resistência aos antibióticos tetraciclina (tet^R), canamicina (kan^R), estreptomicina (sm^R), sulfonamida (su^R) e ampicilina (amp^R), bem como ao mercúrio (hg^R). O segmento determinante de resistência consegue se mover como um grupo de genes de resistência. Tn3 está dentro de Tn4. Cada transpóson pode ser transferido de modo independente. (Dados de S. N. Cohen e J. A. Shapiro, "Transposable Genetic Elements". Direitos autorais 1980 por Scientific American, Inc. Todos os direitos reservados.)

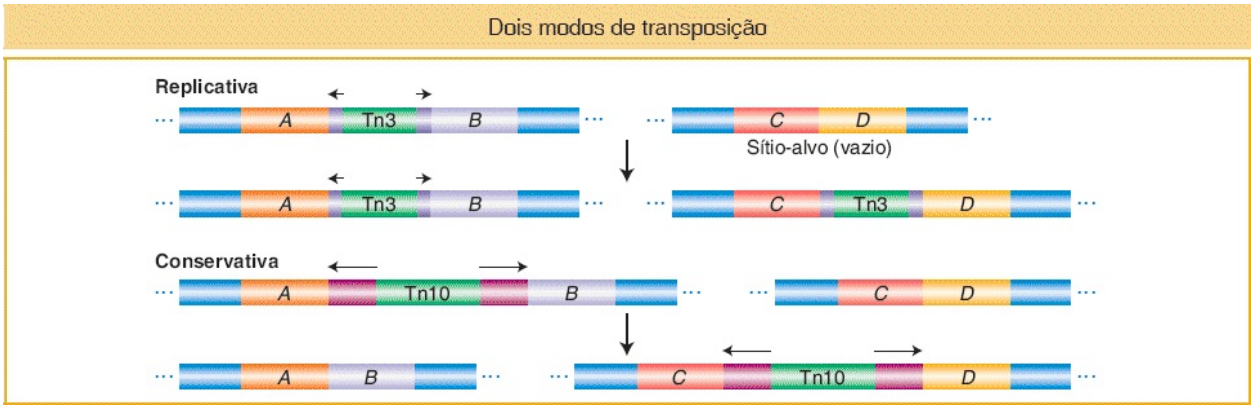


FIGURA 15.8 A transposição de elemento móvel pode ser replicativa ou conservativa. Ver texto para detalhes.

Transposição replicativa de Tn3

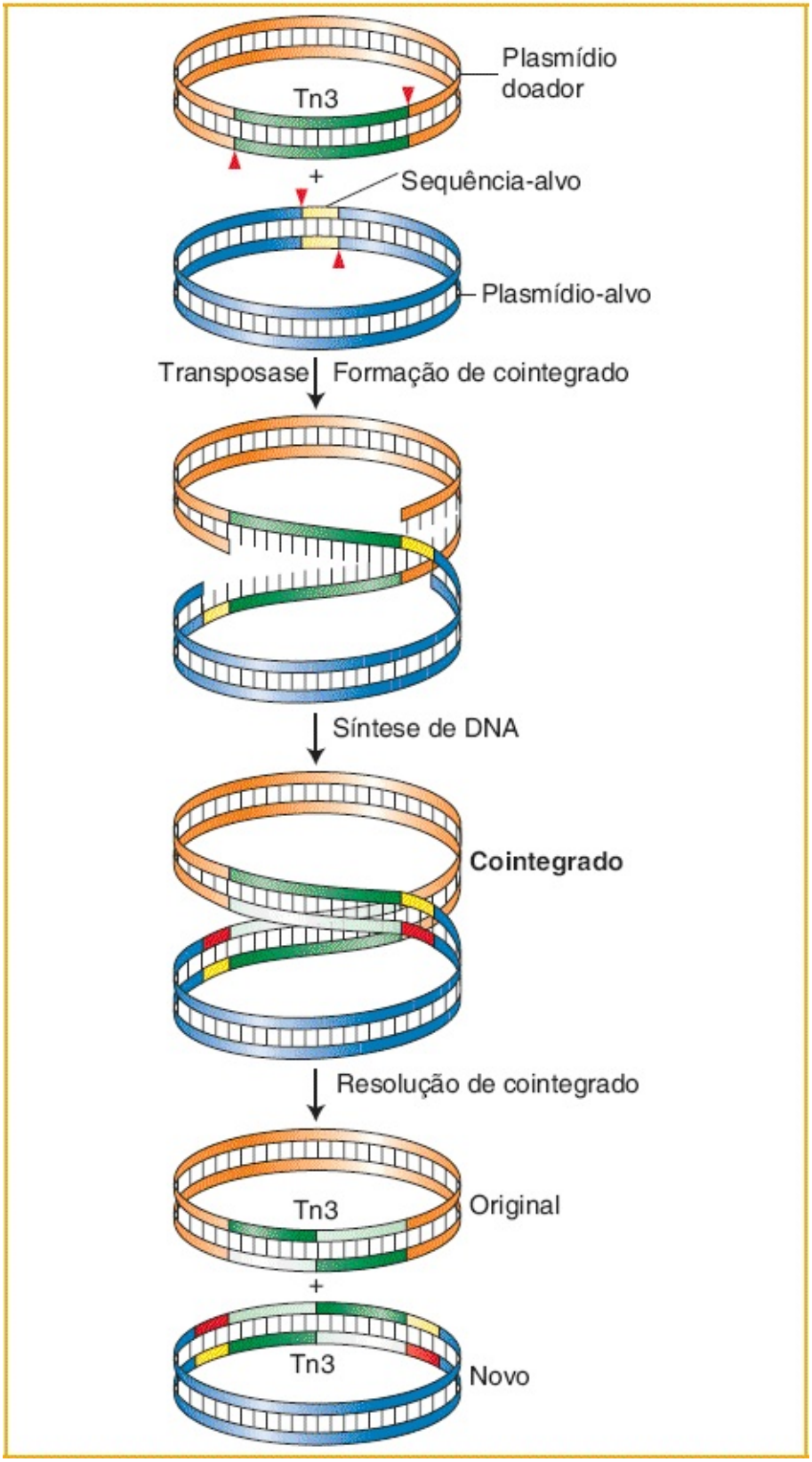


FIGURA 15.9 A transposição replicativa de Tn3 ocorre por meio de um intermediário cointegrado.

Inserção em um novo local. Vimos que a transposase continua a desempenhar um papel importante na inserção. Em uma das primeiras etapas da inserção, a transposase realiza um corte desencontrado no sítio-alvo do DNA (não diferente das quebras desencontradas catalisadas pelas endonucleases de restrição no arcabouço açúcar-fosfato do DNA). A Figura 15.10 demonstra as etapas na inserção de um elemento de transposição genérico. Nesse caso, a transposase realiza um corte desencontrado de cinco pares de bases. O elemento de transposição se insere entre as extremidades cortadas e o maquinário de reparo do DNA hospedeiro (ver Capítulo 16) preenche o intervalo oposto a cada filamento único utilizando as bases das projeções como molde. Agora existem duas sequências duplicadas, cada uma com cinco pares de bases de comprimento, nos sítios das projeções anteriores. Essas sequências são denominadas **duplicação de sítio-alvo.** Praticamente todos os elementos de transposição (tanto em procariotos quanto eucariotos) são flanqueados por uma duplicação de sítio-alvo, indicando que todos utilizam um mecanismo de inserção semelhante àquele demonstrado na Figura 15.10. O que difere é o comprimento da duplicação; um tipo particular de elemento de transposição apresenta um comprimento característico para sua duplicação de sítio-alvo — tão pequeno quanto dois pares de bases para alguns elementos. É importante ter em mente que os elementos de transposição apresentam *repetições invertidas* em suas extremidades e que as repetições invertidas são flanqueadas pela duplicação de sítio-alvo — que é uma *repetição direta*.

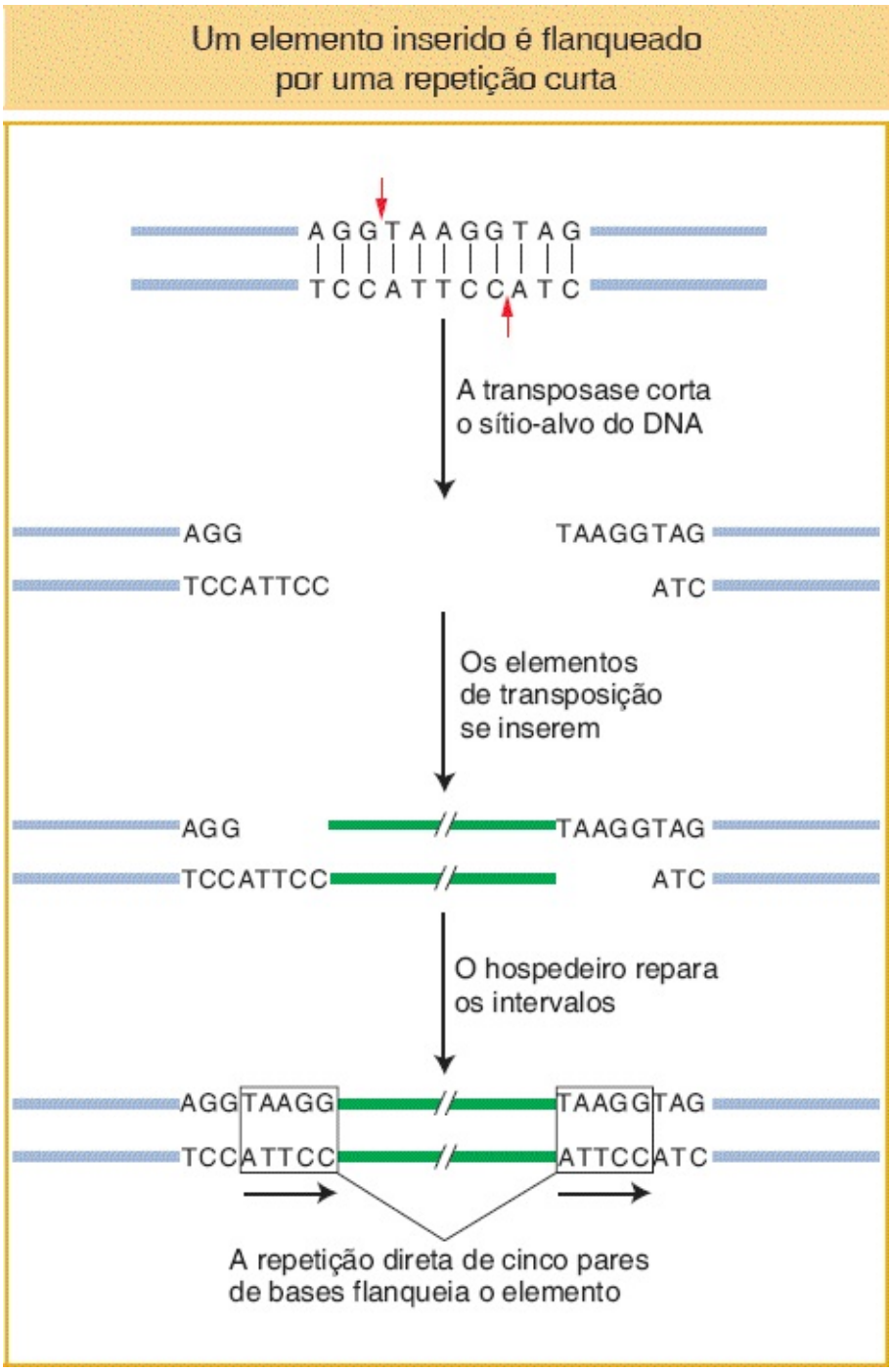


FIGURA 15.10 Uma sequência curta de DNA é duplicada no sítio de inserção do transpóson. O DNA receptor é clivado em sítios desencontrados (está demonstrado um corte desencontrado de 5 pb), levando à produção de duas cópias da sequência de cinco pares de bases que flanqueiam o elemento inserido.

CONCEITO-CHAVE Em procariotos, a transposição ocorre por meio de no

mínimo duas vias diferentes. Alguns elementos de transposição podem replicar uma cópia do elemento em um sítio-alvo, deixando uma cópia para trás no sítio original. Em outros casos, a transposição consiste na excisão direta do elemento e pela sua reinserção dentro em um novo sítio.

15.3 Elementos de transposição em eucariotos

Embora os elementos de transposição tenham sido primeiro descobertos em milho, os primeiros elementos eucarióticos a ser caracterizados no nível molecular foram isolados a partir de genes mutantes de levedura e *Drosophila*. Os elementos de transposição eucarióticos estão situados em duas classes: classe 1 de retrotranspósons e classe 2 de transpósons de DNA. A primeira classe a ser isolada, os retrotranspósons, absolutamente não se assemelha aos elementos IS procarióticos e aos elementos de transposição.

Classe 1 | Retrotranspósons

O laboratório de Gerry Fink estava entre os primeiros a utilizar levedura como um organismo-modelo para estudar a regulação gênica eucariótica. Durante anos, ele e seus colegas isolaram milhares de mutações no gene *HIS4*, que codifica uma das enzimas na via que leva à síntese do aminoácido histidina.

Eles isolaram mais de 1.500 mutantes espontâneos *HIS4* e observaram que dois deles apresentavam um fenótipo mutante instável. Os mutantes instáveis (denominados pseudorrevertentes) apresentavam 1.000 vezes mais probabilidade de reverterem para um fenótipo semelhante ao tipo selvagem do que os outros mutantes *HIS4*. Simbolicamente, dizemos que esses mutantes instáveis reverteram de His^{-} para His^{+} (os tipos selvagens apresentam um sinal de mais sobrescrito, enquanto os mutantes apresentam um sinal de menos sobrescrito). Assim como os mutantes gal^{-} de *E. coli*, observou-se que esses mutantes de levedura abrigam uma grande inserção de DNA no gene *HIS4*. Observou-se que a inserção é muito semelhante àquela de um grupo de elementos de transposição já caracterizados em leveduras, denominados **elementos *Ty***. Existem, de fato, aproximadamente 35

cópias do elemento inserido, denominado *Ty1*, no genoma de levedura.

A clonagem dos elementos desses alelos mutantes levou à descoberta surpreendente de que as inserções absolutamente não se assemelhavam aos elementos IS bacterianos ou transpósons. Em vez disso, elas se assemelhavam a uma classe bem-caracterizada de vírus de animais, denominados retrovírus. Um **retrovírus** é um vírus de RNA unifilamentar que emprega um DNA bifilamentar intermediário para replicação. O RNA é copiado em DNA pela enzima **transcriptase reversa**. O DNA bifilamentar é integrado aos cromossomos hospedeiros, a partir dos quais ele é transcrito para produzir o genoma de RNA viral e as proteínas que formam novas partículas virais. Quando integrada aos cromossomos hospedeiros como DNA bifilamentar, a cópia de DNA bifilamentar do genoma retroviral é denominada **provírus**. O ciclo de vida de um retrovírus típico está demonstrado na Figura 15.11. Alguns retrovírus, tais como o vírus do tumor mamário de camundongos (MMTV) e o vírus do sarcoma de Rous (RSV), são responsáveis pela indução de tumores cancerosos. Em relação ao MMTV, isso ocorre quando ele se insere aleatoriamente no genoma próximo de um gene cuja expressão alterada leva ao câncer.

A Figura 15.12 demonstra a semelhança na estrutura e no conteúdo gênico de um retrovírus e o elemento *Ty1* isolado a partir dos mutantes *HIS4*. Ambos são flanqueados por sequências de **longa repetição terminal (LTR)** que apresentam diversas centenas de pares de bases de comprimento. Os retrovírus codificam no mínimo três proteínas que participam na replicação viral: os produtos dos genes *gag*, *pol* e *env*. A proteína codificada por *gag* apresenta um papel na maturação do genoma de RNA, *pol* codifica a importante transcriptase reversa e *env* codifica a proteína estrutural que circunda o vírus. Essa proteína é necessária para que o vírus saia da célula para infectar outras células. Curiosamente, os elementos *Ty1* apresentam genes relacionados com *gal* e *pol*, mas não com *env*. Essas características levaram à hipótese de que, assim como os retrovírus, os elementos *Ty1* são transcritos em RNA que são copiados em DNA bifilamentar pela transcriptase reversa. Entretanto, contrariamente aos retrovírus, os elementos *Ty1* não conseguem deixar a célula, tendo em vista que não codificam *env*. Em vez disso, as cópias de DNA bifilamentar são inseridas de volta no genoma da mesma

célula. Essas etapas estão diagramadas na Figura 15.13.

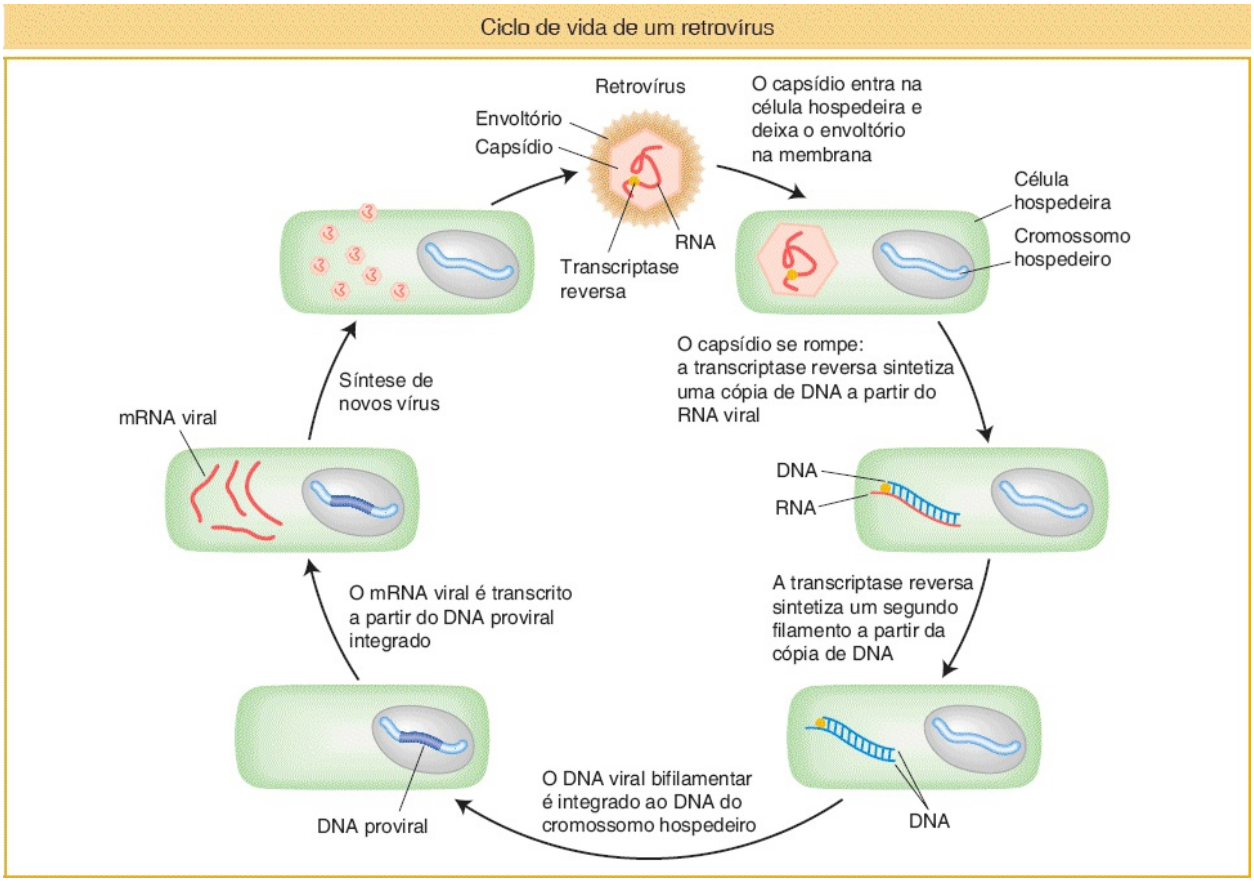


FIGURA 15.11 O genoma de RNA do retrovírus sofre transcrição reversa em DNA bifilamentar dentro da célula hospedeira.

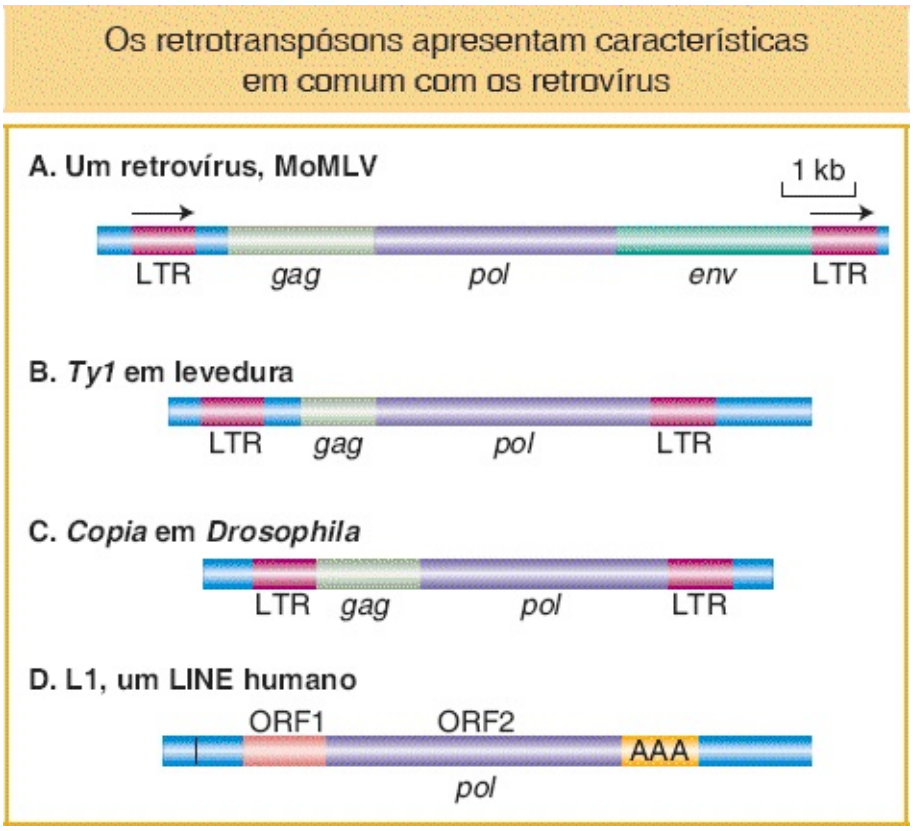


FIGURA 15.12 Comparação estrutural de um retrovírus com retrotranspósons observados em genomas eucarióticos. **A.** Um retrovírus, vírus da leucemia murina Moloney (MoMLV), de camundongos. **B.** Um retrotranspóson, *Ty1*, em levedura. **C.** Um retrotranspóson, *copia*, em *Drosophila*. **D.** Um elemento longo intercalado (LINE) em seres humanos. Abreviações: LTR = Longa repetição terminal; ORF = Matriz de leitura aberta.

Em 1985, David Garfinkel, Jef Boeke e Gerald Fink demonstraram que, assim como os retrovírus, os elementos *Ty* de fato são transpostos por meio de um RNA intermediário. A Figura 15.14 diagrama o desenho experimental desenvolvido por eles. Eles iniciaram com a alteração de um elemento *Ty1* de levedura, clonado em um plasmídio. Primeiramente, perto de uma extremidade de um elemento, eles inseriram um promotor que pode ser ativado pela adição de galactose ao meio. Em segundo lugar, eles introduziram um íntron de outro gene de levedura na região codificadora do transpóson *Ty*.

A adição de galactose aumenta muito a frequência de transposição do elemento *Ty* alterado. Esse aumento na frequência sugere a participação de RNA, tendo em vista que a galactose estimula a transcrição do DNA *Ty* em RNA, com início no

promotor sensível à galactose. O resultado experimental-chave, entretanto, é o destino do DNA *Ty* transposto. Os pesquisadores observaram que o íntron havia sido removido do DNA *Ty* transposto. Tendo em vista que os íntrons são recompostos apenas no período do processamento de RNA (ver Capítulo 8), o DNA *Ty* transposto deve ter sido copiado a partir de um RNA intermediário. A conclusão foi que o RNA é transcrito a partir do elemento *Ty* original e recomposto. O mRNA recomposto sofre transcrição reversa de volta ao DNA bifilamentar, que em seguida é integrado ao cromossomo de levedura. Os elementos de transposição que empregam a transcriptase reversa para transpor por meio de um RNA intermediário são denominados **retrotranspósons**. Eles também são conhecidos como **elementos de transposição da classe 1**. Retrotranspósons tais como *Ty1*, que apresentam *longas repetições terminais* em suas extremidades, são denominados **retrotranspósons LTR** e o mecanismo que eles utilizam para a transposição é denominado "**cópia e colagem**", o que os diferencia do mecanismo de corte e colagem que caracteriza a maior parte dos elementos de transposição de DNA.

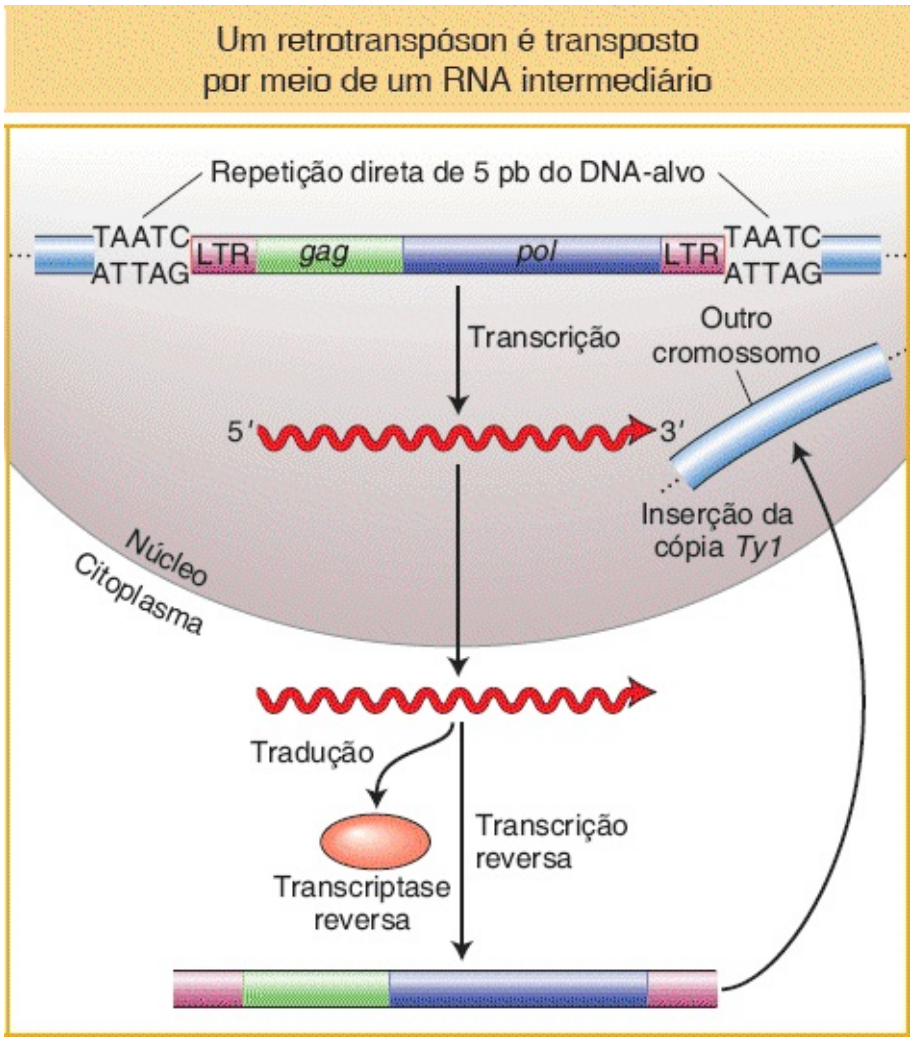


FIGURA 15.13 Um transcrito de RNA do retrotranspóson sofre transcrição reversa em DNA, por meio de uma transcriptase reversa codificada pelo retrotranspóson. A cópia de DNA é inserida em um novo local no genoma.

Também foi demonstrado que diversas mutações espontâneas isoladas ao longo dos anos em *Drosophila* também contêm inserções de retrotranspósons. Os **elementos tipo *copia*** de *Drosophila* são estruturalmente semelhantes aos elementos *Ty1* e aparecem em 10 a 100 posições no genoma de *Drosophila* (ver Figura 15.12 C). Determinadas mutações clássicas de *Drosophila* resultam da inserção de tipo *copia* de outros elementos. Por exemplo, a mutação *white-apricot* (w^a) para cor de olho é causada pela inserção de um elemento da família *copia* no *locus white*. A inserção de retrotranspósons LTR em genes de plantas (incluindo o milho) também demonstrou contribuir para mutações espontâneas

nesse reino.

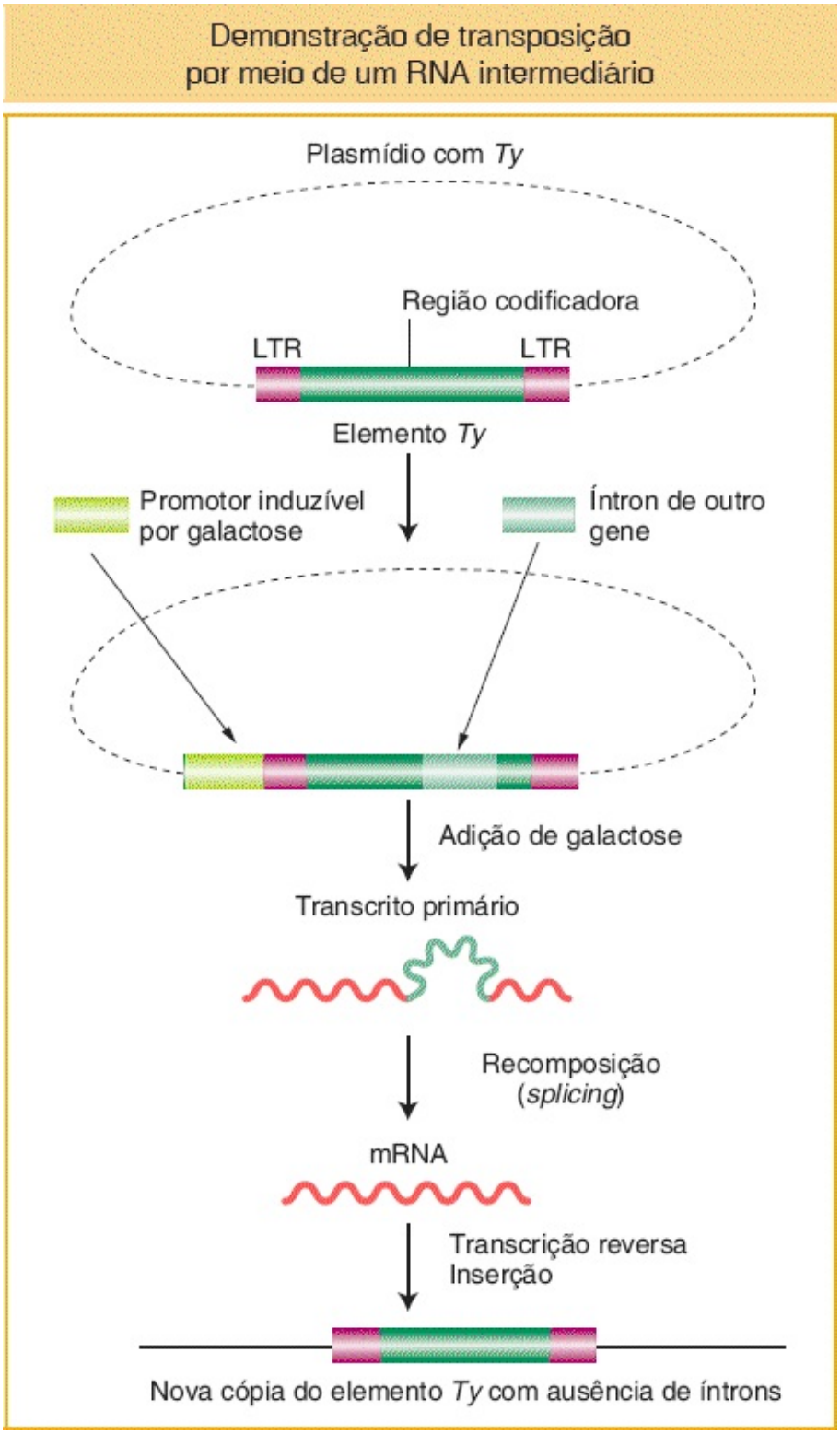


FIGURA 15.14 Um elemento *Ty* é alterado por meio da adição de um íntron e um promotor que pode ser

ativado pela adição de galactose. A sequência do íntron é removida antes da transcrição reversa.

Antes de deixarmos os retrotranspósons (retornaremos a eles posteriormente neste capítulo), existe uma questão que precisa ser respondida. Relembre que o primeiro retrotranspóson LTR foi descoberto em uma linhagem *His*$^{-}$ instável de levedura que com frequência revertia para *His*$^{+}$. Entretanto, acabamos de ver que os retrotranspósons LTR, contrariamente à maior parte dos elementos de transposição de DNA, não são excisados quando transpostos. O que, então, é responsável pelo aumento de aproximadamente 1.000 vezes na frequência de reversão desse alelo quando comparado a outros alelos *His*$^{-}$? A resposta está demonstrada na Figura 15.15, a qual demonstra que o elemento *Ty1* no alelo *His*$^{-}$ está localizado na região promotora do gene *His*, onde ele impede a transcrição gênica. Contrariamente, os revertentes contêm uma única cópia do LTR, denominada **LTR solo**. Essa inserção muito menor não interfere na transcrição do gene *His*. O LTR solo é produto da recombinação entre LTR idênticos, que resulta na deleção do restante do elemento (ver Capítulos 4 e 16 para mais sobre recombinação). Os LTR solo são uma característica muito comum nos genomas de virtualmente todos os eucariotos, indicando a importância desse processo. O genoma de levedura sequenciado contém mais de cinco vezes mais LTR solo do que elementos *Ty1* completos.

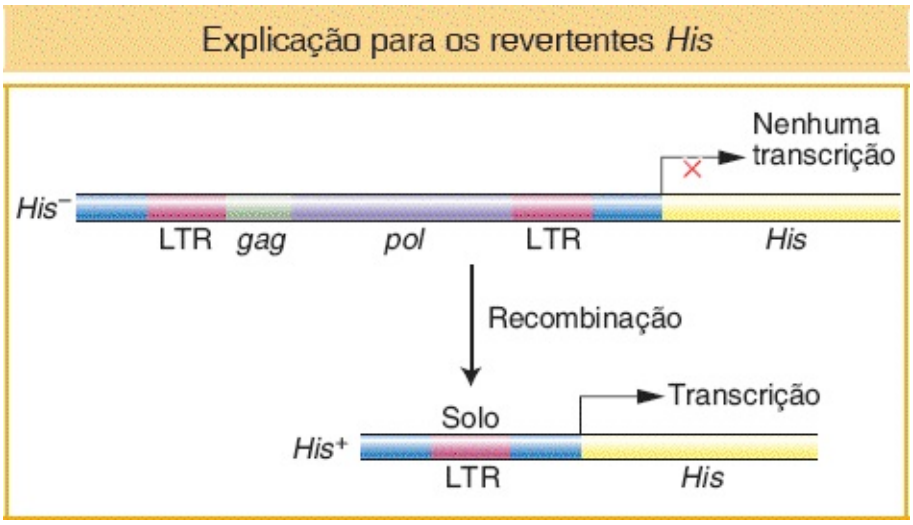


FIGURA 15.15 Revertentes *His*$^{+}$ contêm um LTR solo que resulta da recombinação entre as sequências de

DNA idênticas nos dois LTR do retrotranspóson LTR no promotor *HIS*.

> CONCEITO-CHAVE Os elementos de transposição que são transpostos por RNA intermediários predominam em eucariotos. Os retrotranspósons, também conhecidos como elementos da classe 1, codificam uma transcriptase reversa que produz uma cópia de DNA bifilamentar (de um RNA intermediário) que é capaz de se integrar em uma nova posição no genoma.

Classe 2 | Transpósons de DNA

Alguns elementos móveis observados em eucariotos aparentam ser transpostos por meio de mecanismos semelhantes àqueles em bactérias. Conforme ilustrado na Figura 15.8 em relação aos elementos IS e aos transpósons, o que se insere em uma nova posição no genoma é o próprio elemento ou uma cópia do elemento. Os elementos transpostos dessa maneira são denominados **elementos da classe 2**, ou **transpósons de DNA**. Atualmente sabe-se que os primeiros elementos de transposição descobertos por McClintock em milho são transpósons de DNA. Entretanto, os primeiros transpósons de DNA a serem caracterizados molecularmente foram os elementos *P* em *Drosophila*.

Elementos *P*. De todos os elementos de transposição em *Drosophila*, os mais intrigantes e úteis para os geneticistas são os **elementos *P***. O elemento *P* de tamanho total se assemelha aos transpósons simples de bactérias, uma vez que suas extremidades são repetições invertidas curtas (31 pb), e que codifica uma proteína única — a transposase, que é responsável pela sua mobilização (Figura 15.16). Os elementos *P* variam em tamanho, oscilando de 0,5 a 2,9 kb de comprimento. Essa diferença no tamanho ocorre em virtude da presença de muitos elementos *P* defeituosos a partir dos quais partes do meio do elemento — que codifica o gene da transposase — foram deletadas.

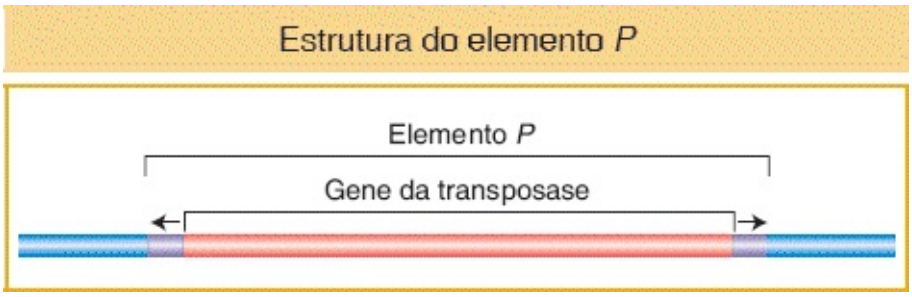


FIGURA 15.16 A análise da sequência de DNA do elemento *P* de 2,9 kb revela um gene que codifica a transposase. Uma repetição invertida perfeita de 31 pb está localizada em cada ponta do elemento.

Os elementos *P* foram descobertos por Margaret Kidwell, que estava estudando a **disgenesia híbrida** — um fenômeno que ocorre quando fêmeas de linhagens de laboratório de *D. melanogaster* são cruzadas com machos derivados de populações naturais. Nesses cruzamentos, diz-se que os estoques de laboratório apresentam um **citotipo M** (tipo celular) e que os estoques naturais apresentam um **citotipo P**. Em um cruzamento de M (fêmea) × P (macho), a progênie demonstra uma diversidade de fenótipos surpreendentes que são manifestados na linhagem germinativa, incluindo esterilidade, uma alta taxa de mutação e uma alta frequência de aberrações cromossômicas e não disjunção (Figura 15.17). Essa progênie híbrida é *disgênica*, ou biologicamente deficiente (daí a expressão *disgenesia híbrida*). Curiosamente, o cruzamento recíproco, P (fêmea) × M (macho), não produz descendência disgênica. Uma observação importante é que uma grande porcentagem das mutações disgenicamente induzidas é instável; ou seja, elas revertem para o tipo selvagem ou para outros alelos mutantes em frequências muito altas. Essa instabilidade em geral é restrita à linhagem germinativa de uma mosca que apresente um citotipo M por meio de um mecanismo explicado posteriormente.

Os mutantes instáveis de *Drosophila* apresentavam semelhanças com os mutantes instáveis de milho caracterizados por McClintock. Os investigadores formularam a hipótese de que as mutações disgênicas são causadas pela inserção de elementos de transposição em genes específicos, tornando-os, assim, inativos. De acordo com essa consideração, a reversão normalmente resultaria da excisão dessas sequências inseridas. Essa hipótese foi testada criticamente por meio do

isolamento de mutações disgênicas instáveis no *locus white* da cor de olho. Observou-se que a maior parte das mutações foi causada pela inserção de um elemento de transposição no gene *white*$^{+}$. Observou-se que o elemento, denominado *elemento P*, está presente em 30 a 50 cópias por genoma nas linhagens P, mas está completamente ausente nas linhagens M.

Por que os elementos *P* não causam problemas em linhagens P? A resposta mais simples é que a transposição do elemento *P* é reprimida em linhagens P. Primeiramente, acreditava-se que a repressão ocorresse em virtude de um repressor proteico que estava nas linhagens P, mas não nas linhagens M. Esse modelo deixou de ser favorecido. Em vez disso, os geneticistas atualmente acreditam que todos os genes da transposase nos elementos *P* são silenciados nas linhagens P. Os genes são ativados na geração F_1, conforme demonstrado na Figura 15.18. O silenciamento gênico foi discutido anteriormente (ver Capítulos 8 e 12) e será revisado no fim deste capítulo. Por algum motivo, a maior parte das linhagens de laboratório não apresenta elementos *P* e consequentemente o mecanismo de silenciamento não é ativado. Em híbridos do cruzamento M (fêmea, sem elementos *P*) × P (machos, com elementos *P*), os elementos *P* no zigoto recém-formado estão em um ambiente livre de silenciamento. Os elementos *P* derivados do genoma masculino agora podem transpor-se pelo genoma diploide, causando uma diversidade de danos na medida em que eles se inserem nos genes e causam mutações. Esses eventos moleculares são expressos como as diversas manifestações de disgenesia híbrida. Por outro lado, os cruzamentos P (fêmea) × M (macho) não resultam em disgenesia presumivelmente em virtude de o citoplasma do zigoto conter os componentes necessários para o silenciamento da transposase do elemento *P*.

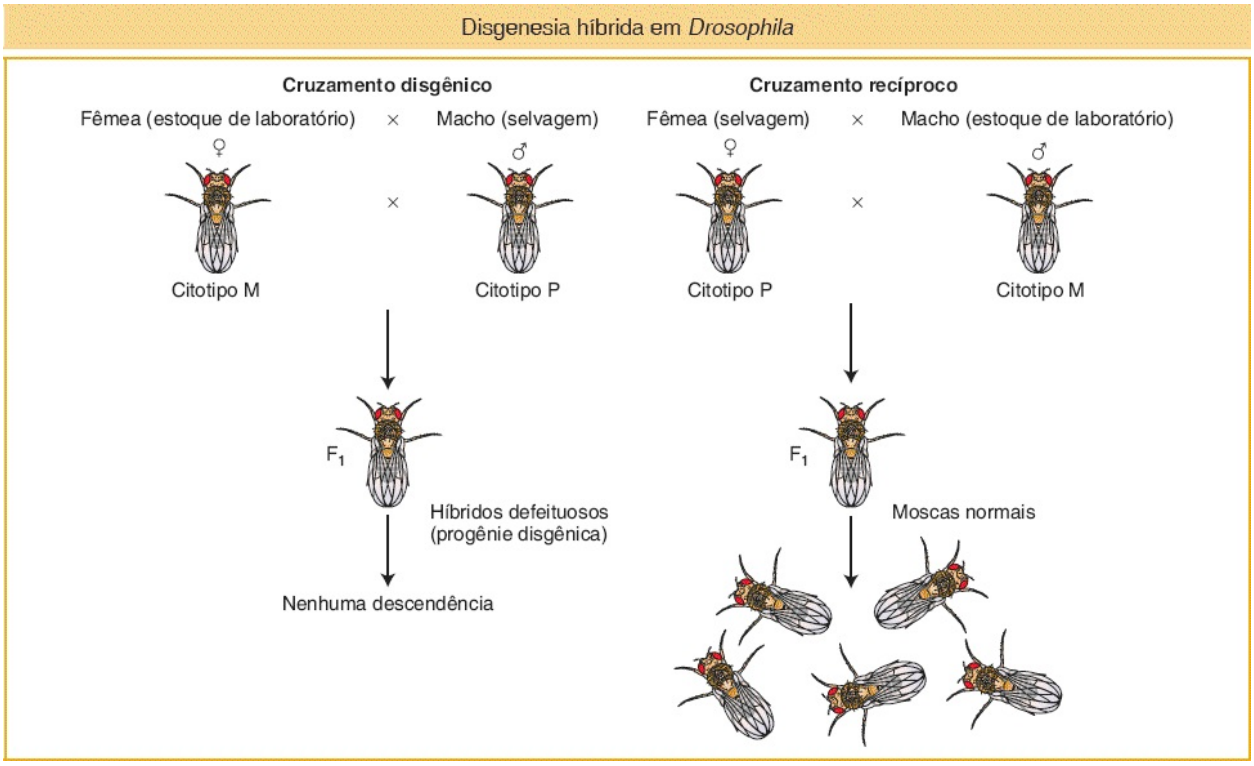


FIGURA 15.17 Na disgenesia híbrida, um cruzamento entre uma fêmea de um estoque de laboratório e um macho selvagem produz progênie defeituosa. Ver texto para os detalhes.

Uma questão intrigante permanece não respondida: por que as linhagens de laboratório não apresentam elementos *P*, enquanto as linhagens no ambiente selvagem apresentam? Uma hipótese é que a maior parte das atuais linhagens de laboratório seja descendente dos isolados originais coletados da natureza por Morgan e seus alunos há quase um século. Em algum ponto entre a captura daquelas linhagens originais e as atuais, os elementos *P* se difundiram pelas populações naturais, mas não pelas linhagens de laboratório. Essa diferença não foi notada até que linhagens selvagens fossem novamente capturadas e cruzadas com linhagens de laboratório.

Embora não saibamos exatamente como os elementos *P* foram disseminados, está claro que os elementos de transposição podem se disseminar rapidamente a partir de alguns membros de uma população. Nesse sentido, a disseminação dos elementos *P* se assemelha à disseminação dos transpósons que carreiam genes de resistência para populações bacterianas anteriormente suscetíveis.

Revisão dos elementos de transposição em milho. Embora o agente causal responsável pelos mutantes instáveis tenha sido primeiramente demonstrado geneticamente como os elementos de transposição em milho, isso ocorreu quase 50 anos antes que os elementos *Ac* e *Ds* de milho fossem isolados e que se demonstrasse que eles estão relacionados com os transpósons de DNA em bactérias e em outros eucariotos. Assim como o elemento *P* de *Drosophila*, *Ac* apresenta repetições terminais invertidas e codifica uma proteína única, a transposase. O elemento *Ds* não autônomo não codifica a transposase e, portanto, não consegue realizar a transposição por conta própria. Quando *Ac* está presente no genoma, sua transposase consegue se ligar a ambas as extremidades dos elementos *Ac* ou *Ds* e promover a sua transposição (Figura 15.19).

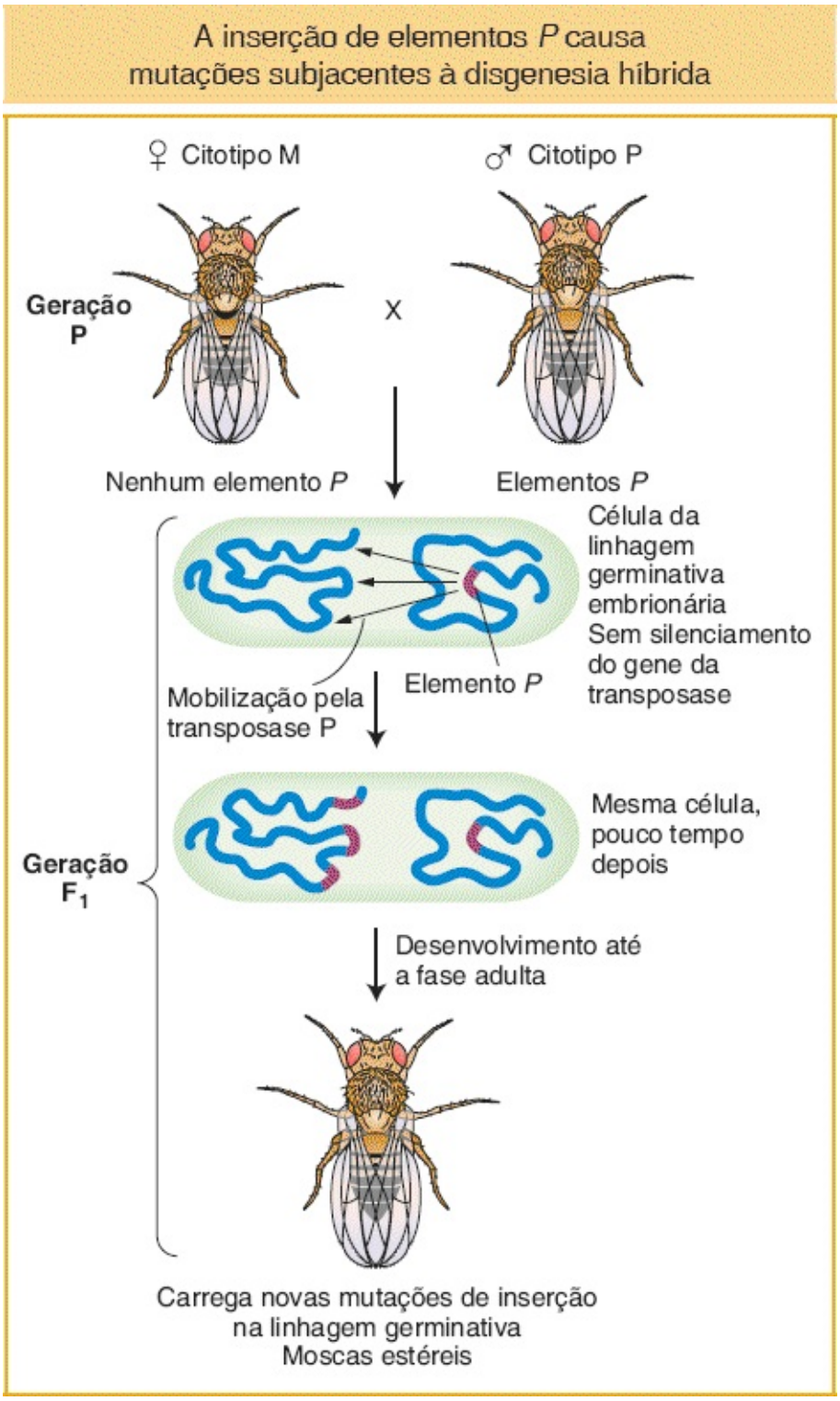


FIGURA 15.18 Eventos moleculares subjacentes à disgenesia híbrida. Cruzamentos de macho de *Drosophila* que contém a transposase P com fêmea de *Drosophila* que não apresenta elementos *P* funcionais produzem mutações na linhagem germinativa da progênie F_1 causadas por inserções do elemento *P*. Os elementos *P* são capazes de se mover, causando mutações, tendo em vista que o zigoto não silencia o gene da

transposase.

Conforme observado anteriormente no capítulo, *Ac* e *Ds* são membros de uma única família de transpósons e existem outras famílias de elementos de transposição em milho. Cada família contém um elemento autônomo que codifica uma transposase capaz de movimentar elementos na mesma família, mas não de movimentar elementos em outras famílias, tendo em vista que a transposase consegue se ligar apenas às extremidades de membros da família.

Embora alguns organismos tais como leveduras não apresentem transpósons de DNA, elementos estruturalmente semelhantes aos elementos *P* e *Ac* foram isolados de muitas espécies de plantas e animais.

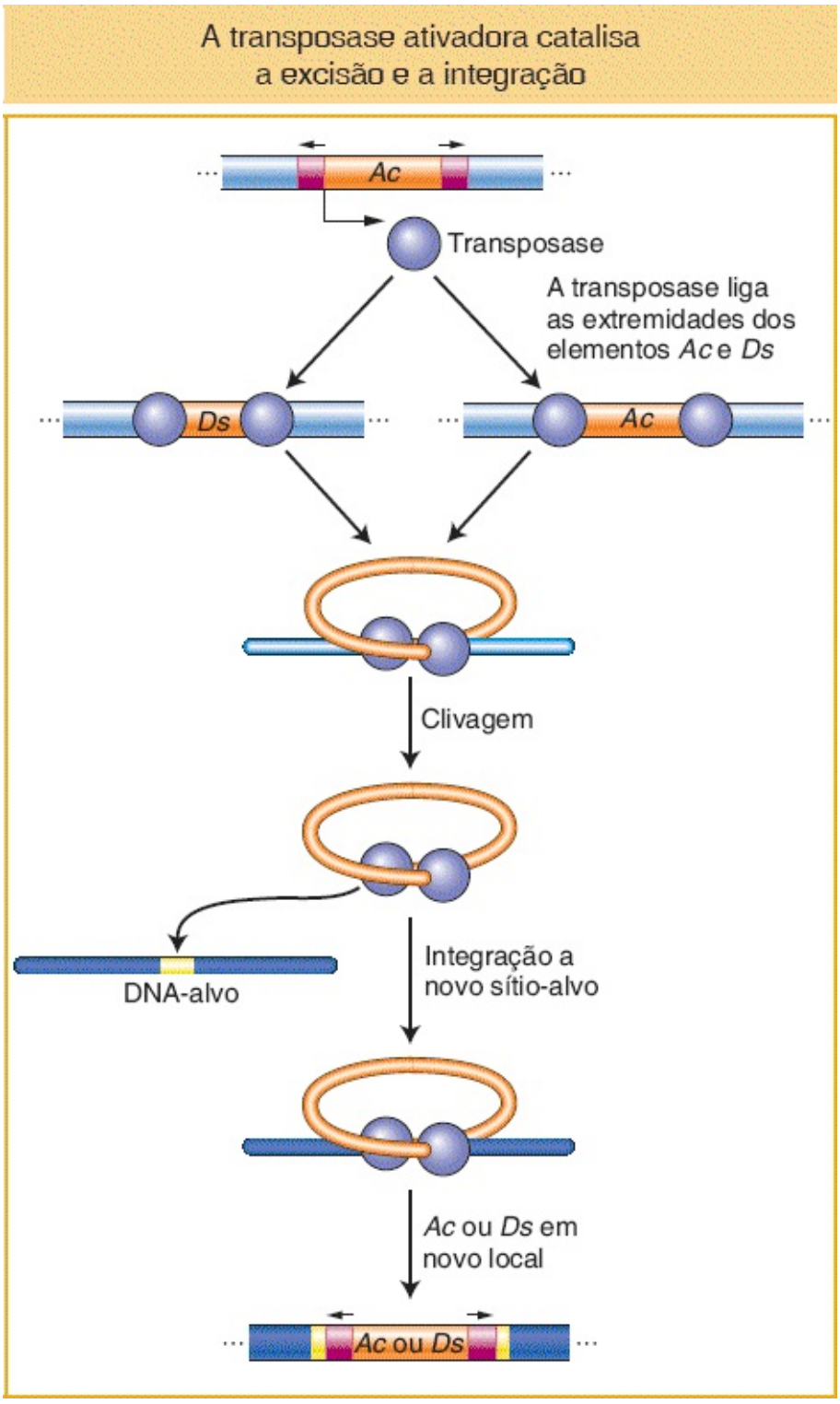


FIGURA 15.19 O elemento *Ac* em milho codifica uma transposase que liga as suas próprias extremidades ou aquelas de um elemento *Ds*, excisando o elemento, clivando o sítio-alvo e possibilitando que o elemento se insira em algum outro local no genoma.

CONCEITO-CHAVE Os primeiros elementos de transposição conhecidos em milho são transpósons de DNA que se assemelham estruturalmente aos transpósons de DNA em bactérias e outros eucariotos. Os transpósons de DNA codificam uma transposase que corta o transpóson do cromossomo e catalisa a sua reinserção em outros locais cromossômicos.

Utilidade dos transpósons de DNA para a descoberta de genes

Além de seu interesse como um fenômeno genético, os transpósons de DNA se tornaram ferramentas importantes utilizadas pelos geneticistas que trabalham com uma diversidade de organismos. Sua mobilidade foi explorada para marcar genes para clonagem e para inserir transgenes. O elemento *P* em *Drosophila* fornece um dos melhores exemplos de como os geneticistas exploram as propriedades dos elementos de transposição em eucariotos.

Utilização de elementos P para marcar genes para clonagem. Os elementos *P* podem ser utilizados para criar mutações por meio de inserção, marcar a posição dos genes e facilitar a clonagem dos genes. Os elementos *P* inseridos nos genes *in vivo* rompem os genes aleatoriamente, criando mutantes com fenótipos diferentes. Moscas-das-frutas com fenótipos mutantes interessantes podem ser selecionadas para clonagem do gene mutante, que é marcado pela presença do elemento *P*, um método denominado **marcação de transpóson**. Após a clonagem do gene interrompido, os fragmentos do alelo mutante podem ser utilizados como uma sonda para isolar o gene do tipo selvagem.

Utilização de elementos *P* para inserir genes. Gerald Rubin e Allan Spradling demonstraram que o DNA do elemento *P* pode ser um veículo eficaz para a transferência dos genes doadores para a linhagem germinativa de uma mosca receptora. Eles planejaram o procedimento experimental a seguir (Figura 15.20). Suponha que o objetivo é transferir o alelo ry^+, que confere uma cor de olho característica, para o genoma da mosca. O genótipo receptor é homozigoto para a mutação *rosy* (ry^-). A partir dessa linhagem, os embriões são coletados ao

término de aproximadamente nove divisões nucleares. Nesse estágio, o embrião é uma célula multinucleada e os núcleos que se destinam à formação das células germinativas estão agrupados em uma extremidade. (Os elementos *P* mobilizam apenas nas células da linhagem germinativa.) Dois tipos de DNA são injetados em embriões desse tipo. O primeiro é um plasmídio bacteriano que carreia um elemento *P* defeituoso no qual o gene ry^+ foi inserido. O elemento *P* defeituoso se assemelha ao elemento *Ds* de milho, pois não codifica a transposase, mas ainda apresenta as extremidades que ligam a transposase e possibilitam a transposição. Esse elemento deletado não é capaz de se transpor e, assim, conforme mencionado anteriormente, também é injetado um plasmídio auxiliar que codifica a transposase, porém sem as repetições terminais (de modo que ele não consegue se transpor). As moscas que se desenvolvem a partir desses embriões ainda são fenotipicamente mutantes *rosy*, mas sua descendência inclui uma grande proporção de moscas ry^+. A hibridização *in situ* confirmou que o gene ry^+, juntamente com o elemento *P* deletado, foi inserido em um de diversos locais cromossômicos distintos. Nenhum apareceu exatamente no *locus* normal do gene *rosy*. Observou-se que esses novos genes ry^+ são herdados de modo mendeliano estável.

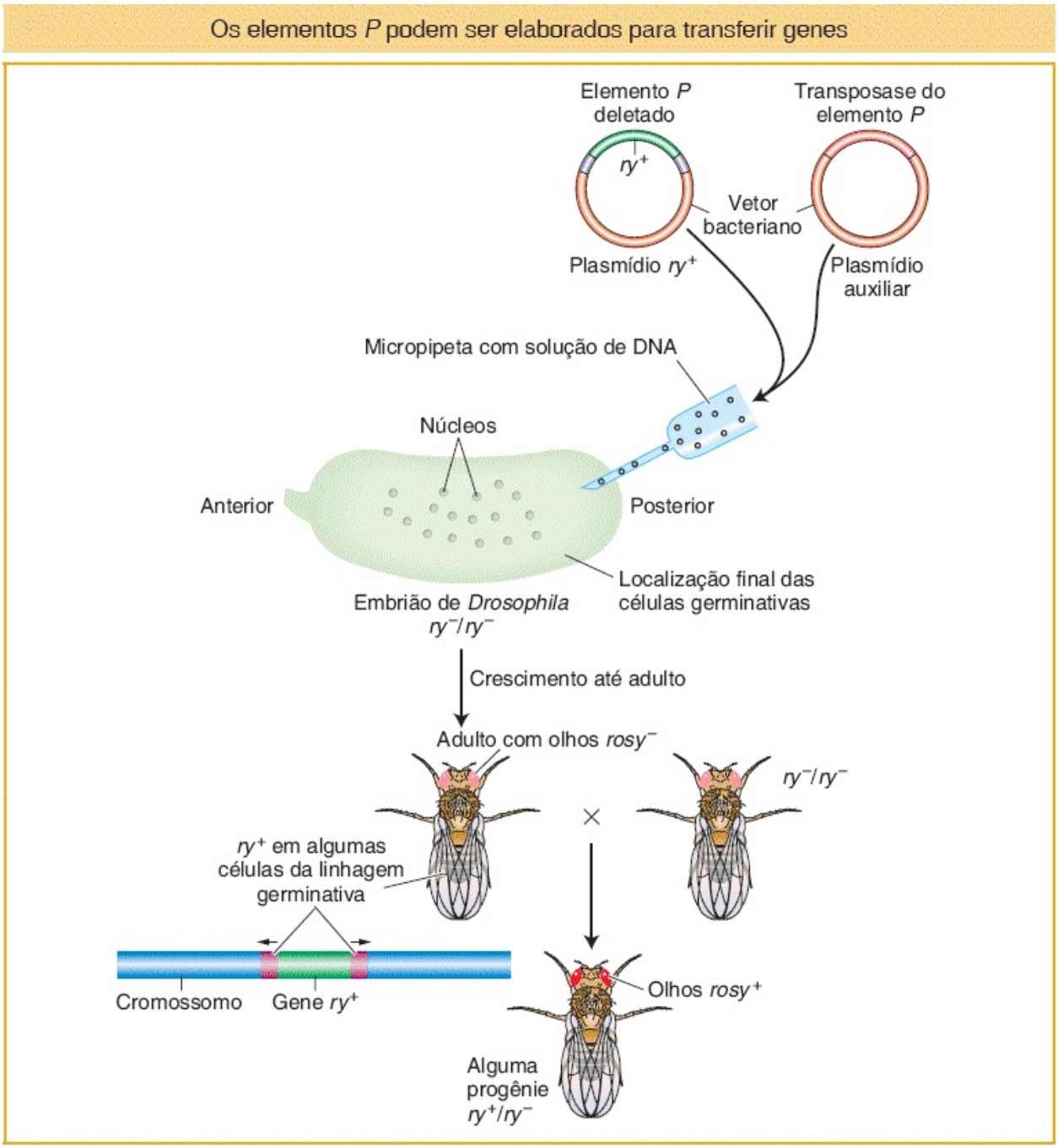


FIGURA 15.20 Transferência gênica mediada por elemento *P* em *Drosophila*. O gene de cor de olho $rosy^+$ (ry^+) é inserido em um elemento *P* deletado carreado em um vetor bacteriano. Ao mesmo tempo, é utilizado um plasmídio auxiliar que contém um elemento *P* intacto. Ambos são injetados em um embrião ry^-, no qual o ry^+ é transposto com o elemento *P* para os cromossomos das células da linhagem germinativa.

Tendo em vista que o elemento *P* pode ser transposto apenas em *Drosophila*, tais aplicações são restritas a essas moscas. Contrariamente, o elemento *Ac* de

milho é capaz de se transpor após sua introdução nos genomas de muitas espécies de plantas, incluindo *Arabidopsis*, alface, cenoura, arroz e cevada. Assim como os elementos *P*, *Ac* foi modificado por geneticistas para a utilização no isolamento de genes por meio da marcação de transpóson. Desse modo, o *Ac*, o primeiro elemento de transposição descoberto por Barbara McClintock, atua como uma importante ferramenta utilizada por geneticistas de plantas há mais de 50 anos.

> CONCEITO-CHAVE Os transpósons de DNA têm sido modificados e utilizados por cientistas de dois modos importantes: (1) para produzir mutantes que possam ser molecularmente identificados pela presença de uma marcação de transpóson e (2) como vetores que podem introduzir genes exógenos em um cromossomo.

15.4 O genoma dinâmico | Mais elementos de transposição do que jamais se imaginou

Os elementos de transposição foram primeiro descobertos com a utilização de abordagens genéticas. Nesses estudos, os elementos tornaram sua presença conhecida quando foram transpostos para um gene ou em sítios de quebra ou rearranjo cromossômico. Após o DNA de elementos de transposição ter sido isolado de mutações instáveis, os cientistas puderam utilizá-lo como sondas moleculares para determinar se havia mais cópias relacionadas no genoma. Em todos os casos, diversas cópias — e em alguns casos várias centenas de cópias — do elemento sempre estavam presentes no genoma.

Os cientistas ponderaram a respeito da prevalência dos elementos de transposição nos genomas. Haveria outros elementos de transposição no genoma que permaneciam desconhecidos em virtude de não terem causado uma mutação que pudesse ser estudada em laboratório? Haveria elementos de transposição na vasta maioria dos organismos que não eram passíveis de análise genética?

Perguntando de outro modo, os organismos sem mutações induzidas por elementos de transposição, ainda assim, apresentam elementos de transposição em seus genomas? Essas questões são reminiscentes da experimentação do pensamento filosófico: “Se uma árvore cai na floresta e ninguém está perto para ouvir, ela produz som?”

Grandes genomas são predominantemente elementos de transposição

Muito antes do advento dos projetos de sequenciamento de DNA, cientistas que utilizavam uma diversidade de técnicas bioquímicas descobriram que o conteúdo de DNA (denominado **valor *C***) variava dramaticamente em eucariotos e não estava correlacionado com a complexidade biológica. Por exemplo, os genomas de salamandras são 20 vezes maiores que o genoma humano, enquanto o genoma da cevada é mais de 10 vezes maior que o genoma do arroz, uma gramínea correlata. A ausência de correlação entre o tamanho do genoma e a complexidade biológica de um organismo é conhecida como **paradoxo do valor *C***.

A cevada e o arroz são, ambos, gramíneas de cereais e, como tal, o conteúdo do seu genoma deveria ser semelhante. Entretanto, se os genes são um componente relativamente constante dos genomas de organismos multicelulares, o que é responsável pelo DNA adicional nos genomas maiores? Com base nos resultados de experimentos adicionais, os cientistas foram capazes de determinar que as sequências de DNA que são repetidas milhares, até mesmo centenas de milhares de vezes, compõem uma grande fração dos genomas eucarióticos e que alguns genomas contêm muito mais DNA repetitivo do que outros.

Graças a muitos projetos recentes para sequenciar os genomas de uma ampla variedade de organismos (incluindo *Drosophila*, seres humanos, camundongos, *Arabidopsis* e arroz), atualmente sabemos que existem muitas classes de sequências repetitivas nos genomas de organismos superiores e que algumas são semelhantes aos transpósons de DNA e aos retrotranspósons mostrados como responsáveis por mutações em plantas, leveduras e insetos. Mais notavelmente, essas sequências compõem a maior parte do DNA nos genomas de eucariotos multicelulares.

Em vez de se correlacionar com o conteúdo gênico, o tamanho do genoma com

frequência se correlaciona com a quantidade de DNA no genoma que é derivada de elementos de transposição. Organismos com genomas grandes apresentam muitas sequências que se assemelham aos elementos de transposição, enquanto organismos com genomas pequenos apresentam muito menos. Dois exemplos, um do genoma humano e o outro de uma comparação dos genomas de gramíneas, ilustram esse ponto. As características estruturais dos elementos de transposição que são observados nos genomas humanos estão resumidas na Figura 15.21 e será feita referência a elas na próxima seção.

> CONCEITO-CHAVE O paradoxo do valor *C* é a ausência de correlação entre o tamanho do genoma e a complexidade biológica. Os genes compõem apenas uma pequena proporção dos genomas de organismos multicelulares. O tamanho do genoma normalmente corresponde à quantidade de sequências de elementos de transposição, em vez de ao conteúdo gênico.

Tipos de elementos de transposição no genoma humano

Elemento	Transposição	Estrutura	Tamanho	Número de cópias	Fração do genoma
LINE	Autônoma	ORF1 ORF2 *(pol)* AAA	1 a 5 kb	20.000 a 40.000	21%
SINE	Não autônoma	AAA	100 a 300 pb	1.500.000	13%
Transpósons de DNA	Autônoma	transposase	2 a 3 kb	300.000	3%
	Não autônoma		80 a 3.000 pb		

FIGURA 15.21 Diversas classes gerais de elementos de transposição são observadas no genoma humano. (Dados de *Nature* 409, 880 [15 de fevereiro de 2001], "Initial Sequencing and Analysis of the Human Genome", The International Human Genome Sequencing Consortium.)

Elementos de transposição no genoma humano

Quase metade do genoma humano é derivada de elementos de transposição. A maioria desses elementos de transposição é de dois tipos de retrotranspósons, denominados **elementos longos intercalados**, ou **LINE**, e **elementos curtos**

intercalados, ou **SINE** (ver Figura 15.21). Os LINE se movem como um retrotranspóson com o auxílio de uma transcriptase reversa codificada por um elemento, mas não apresentam algumas características estruturais de elementos do tipo retrovírus, incluindo LTR (ver Figura 15.12 D). Os SINE podem ser mais bem-descritos como LINE não autônomos, tendo em vista que eles apresentam as características estruturais de LINE, mas não codificam a sua própria transcriptase reversa. Presumivelmente, eles são mobilizados por enzimas transcriptases reversas que são codificadas por LINE presentes no genoma.

O SINE mais abundante em seres humanos é denominado ***Alu***, em virtude de conter um sítio-alvo para a enzima de restrição Alu. O genoma humano contém mais de 1 milhão de sequências *Alu* inteiras e parciais, dispersas entre os genes e dentro de íntrons. Essas sequências *Alu* compõem mais de 10% do genoma humano. A sequência *Alu* total apresenta aproximadamente 200 nucleotídios de comprimento e apresenta uma semelhança extraordinária com o RNA 7SL, um RNA que é parte de um complexo por meio do qual polipeptídios recém-sintetizados são secretados pelo retículo endoplasmático. Presumivelmente, as sequências *Alu* foram originadas como transcritos reversos dessas moléculas de RNA.

Existe aproximadamente 20 vezes mais DNA no genoma humano derivado de elementos de transposição do que DNA codificador de todas as proteínas humanas. A Figura 15.22 ilustra o número e a diversidade de elementos de transposição presentes no genoma humano, utilizando como exemplo as posições de *Alu* individuais, outros SINE e LINE na vizinhança de um gene humano típico.

O genoma humano parece ser típico de um organismo multicelular na abundância e na distribuição de elementos de transposição. Portanto, uma questão óbvia é, Como as plantas e os animais sobrevivem com tantas inserções nos genes e tanto DNA móvel no genoma? Primeiramente, em relação à função dos genes, todos os elementos demonstrados na Figura 15.22 estão inseridos em íntrons. Portanto, o mRNA produzido por esse gene não incluirá quaisquer sequências de elementos de transposição, tendo em vista que eles terão sido removidos do pré-mRNA com o íntron adjacente. Presumivelmente, os elementos de transposição se inserem tanto em éxons quanto em íntrons, mas apenas as inserções em íntrons

permanecerão na população, tendo em vista que apresentam menos probabilidade de causar uma mutação deletéria. Diz-se que as inserções em éxons estão sujeitas à **seleção negativa**. Em segundo lugar, os seres humanos, assim como todos os outros organismos multicelulares, conseguem sobreviver com tanto DNA móvel no genoma em virtude de a vasta maioria ser inativa e não conseguir se mover ou aumentar o número de cópias. A maior parte das sequências de elementos de transposição em um genoma é de remanescentes que acumularam mutações inativadoras durante o período evolutivo. Outros ainda são capazes de movimentação, mas se tornam inativos pelos mecanismos reguladores do hospedeiro (ver Seção 15.5). Entretanto, existem alguns LINE e *Alu* ativos que conseguiram escapar do controle do hospedeiro e inseriram-se em genes importantes, causando diversas doenças humanas. Três inserções de LINE em separado romperam o gene do fator VIII, causando a hemofilia A. No mínimo 11 inserções *Alu* em genes humanos demonstraram causar diversas doenças, incluindo a hemofilia B (no gene do fator IX), a neurofibromatose (no gene *NF1*) e o câncer de mama (no gene *BRCA2*).

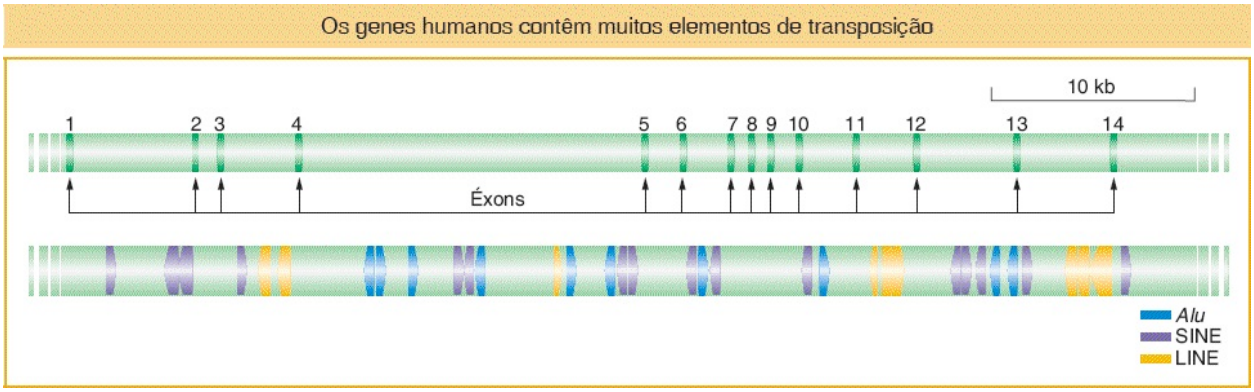


FIGURA 15.22 Diversos elementos repetitivos são observados no gene humano (*HGO*) que codifica a homogentisato 1,2-dioxigenase, a enzima cuja deficiência causa alcaptonúria. A fileira superior diagrama as posições dos éxons de *HGO*. As localizações de *Alu* (*azul*), outros SINE (*roxo*) e LINE (*amarelo*) na sequência de *HGO* estão indicadas na fileira inferior. (Dados de B. Granadino, D. Beltrán-Valero de Bernabé, J. M. Fernández-Cañón, M. A. Peñalva e S. Rodríguez de Córdoba, "The Human Homogentisate 1,2-dioxygenase (HGO) gene", *Genomics* 43, 1997, 115.)

A frequência geral de mutação espontânea em virtude da inserção de elementos da classe 2 em seres humanos é razoavelmente baixa, responsável por menos de

0,2% (1 em 500) de todas as mutações espontâneas caracterizadas. Surpreendentemente, as inserções de retrotranspósons são responsáveis por aproximadamente 10% das mutações espontâneas em outro mamífero, o camundongo. O aumento de aproximadamente 50 vezes nesse tipo de mutação no camundongo muito provavelmente corresponde à atividade muito mais alta desses elementos no genoma de camundongo do que no genoma humano.

> CONCEITO-CHAVE Os elementos de transposição compõem a maior fração do genoma humano, com LINE e SINE sendo os mais abundantes. A maioria dos elementos de transposição é resquício ancestral que não consegue mais se mover ou aumentar seu número de cópias. Alguns poucos elementos permanecem ativos e sua movimentação para os genes pode causar doenças.

As gramíneas | Retrotranspósons LTR desenvolvem-se em grandes genomas

Conforme já mencionado, o paradoxo do valor *C* é a ausência de correlação entre o tamanho do genoma e a complexidade biológica. Como os organismos podem apresentar conteúdo gênico muito semelhante, mas diferir dramaticamente no tamanho de seus genomas? Essa situação foi investigada nas gramíneas de cereais. As diferenças nos tamanhos dos genomas dessas gramíneas demonstraram estar correlacionadas primariamente com o número de uma classe de elementos, os retrotranspósons LTR. Os cereais são parentes evolutivos que tiveram origem a partir de um ancestral comum nos últimos 70 milhões de anos. Como tal, seus genomas ainda são muito semelhantes em relação ao conteúdo gênico e à organização (denominada **sintenia**; ver Capítulo 14) e regiões podem ser comparadas diretamente. Essas comparações revelam que genes ligados no pequeno genoma de arroz estão fisicamente mais próximos do que os mesmos genes nos genomas maiores de milho e cevada. Nos genomas de milho e cevada, os genes são separados por grandes grupos de retrotranspósons (Figura 15.23).

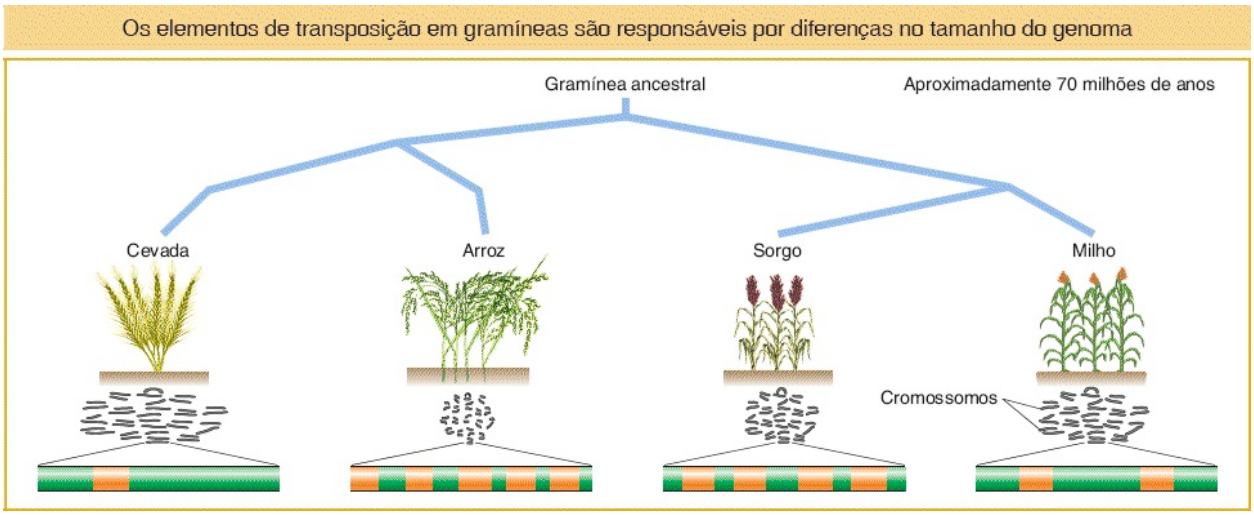


FIGURA 15.23 As gramíneas, incluindo cevada, arroz, sorgo e milho, divergiram a partir de um ancestral comum há aproximadamente 70 milhões de anos. Desde aquela época, os elementos de transposição acumularam-se em níveis diferentes em cada espécie. Os cromossomos são maiores em milho e cevada, cujos genomas contêm grandes quantidades de retrotranspósons LTR. A cor *verde* no genoma parcial na parte inferior representa um aglomerado de transpósons, enquanto *laranja* representa os genes.

Abrigos seguros

A abundância de elementos de transposição nos genomas de organismos multicelulares levou alguns investigadores a postular que elementos de transposição bem-sucedidos (aqueles que são capazes de alcançar números muito altos de cópias) desenvolveram mecanismos para prevenir danos aos seus hospedeiros por meio da não inserção em genes hospedeiros. Em vez disso, os elementos de transposição bem-sucedidos se inserem dentro dos assim denominados **abrigos seguros** no genoma. Em relação às gramíneas, um abrigo seguro para novas inserções aparenta ser em outros retrotranspósons. Outro abrigo seguro é a heterocromatina dos centrômeros, na qual existem muito poucos genes, mas muito DNA repetitivo (ver Capítulo 12 para mais sobre a heterocromatina). Muitas classes de elementos de transposição em espécies de plantas e de animais tendem a inserir-se na heterocromatina centromérica.

Abrigos seguros em genomas pequenos | Inserções dirigidas. Contrariamente aos genomas de eucariotos multicelulares, o genoma de leveduras unicelulares é muito compacto, com genes próximos e muito poucos íntrons. Com quase 70% de

seu genoma como éxons, existe uma alta probabilidade de que novas inserções de elementos de transposição interrompam uma sequência codificadora. Ainda assim, conforme vimos anteriormente neste capítulo, o genoma de leveduras ampara uma coleção de retrotranspósons LTR denominados elementos *Ty*.

Como os elementos de transposição são capazes de se difundir para novos sítios nos genomas com poucos abrigos seguros? Investigadores identificaram centenas de elementos *Ty* no genoma de leveduras sequenciado e determinaram que eles não estão distribuídos aleatoriamente. Em vez disso, cada família de elementos *Ty* insere-se em uma região genômica em particular. Por exemplo, a família *Ty3* insere-se quase exclusivamente próximo dos genes de tRNA, embora não nesses genes, em sítios onde eles não interferem com a produção de tRNA e, presumivelmente, não prejudicam seus hospedeiros. Os elementos *Ty* desenvolveram um mecanismo que possibilita a sua inserção em regiões do genoma em particular: as proteínas *Ty* necessárias para integração interagem com proteínas específicas de levedura ligadas ao DNA genômico. As proteínas *Ty3*, por exemplo, reconhecem e se ligam às subunidades do complexo da RNA polimerase que foram reunidas em promotores de tRNA (Figura 15.24 A).

A capacidade de alguns transpósons de se inserirem preferencialmente em determinadas sequências ou regiões genômicas é denominada **direcionamento**. Um exemplo surpreendente de direcionamento é ilustrado pelos elementos *R1* e *R2* de artrópodes, incluindo a *Drosophila*. *R1* e *R2* são LINE (ver Figura 15.21) que se inserem apenas em genes que produzem RNA ribossômico. Em artrópodes, diversas centenas de genes de rRNA estão organizadas em arranjos em *tandem* (Figura 15.24 B). Com tantos genes codificando o mesmo produto, o hospedeiro tolera a inserção em um subconjunto. Entretanto, demonstrou-se que muitas inserções de *R1* e *R2* diminuem a viabilidade do inseto, presumivelmente por meio de interferência na montagem do ribossomo.

Revisão da terapia gênica. Este capítulo foi iniciado com uma descrição de um distúrbio genético recessivo denominado SCID (doença por imunodeficiência combinada grave). Os sistemas imunológicos de pessoas afetadas pela SCID estão gravemente comprometidos em virtude de uma mutação em um gene que

codifica a enzima adenosina desaminase. Para corrigir esse defeito genético, células da medula óssea de pacientes com SCID foram coletadas e tratadas com um vetor retrovírus contendo um gene ADA normal. Em seguida, as células transformadas foram infundidas de volta nos pacientes. Os sistemas imunólogos da maior parte dos pacientes demonstraram melhora significativa. Entretanto, a terapia apresentou um efeito colateral muito sério: dois dos pacientes desenvolveram leucemia. Em ambos os pacientes, o vetor retroviral inseriu-se (integrou-se) próximo de um gene celular cuja expressão aberrante está associada à leucemia. Um provável cenário é que aquela inserção do vetor retroviral próximo ao gene celular alterou a sua expressão e, indiretamente, causou a leucemia.

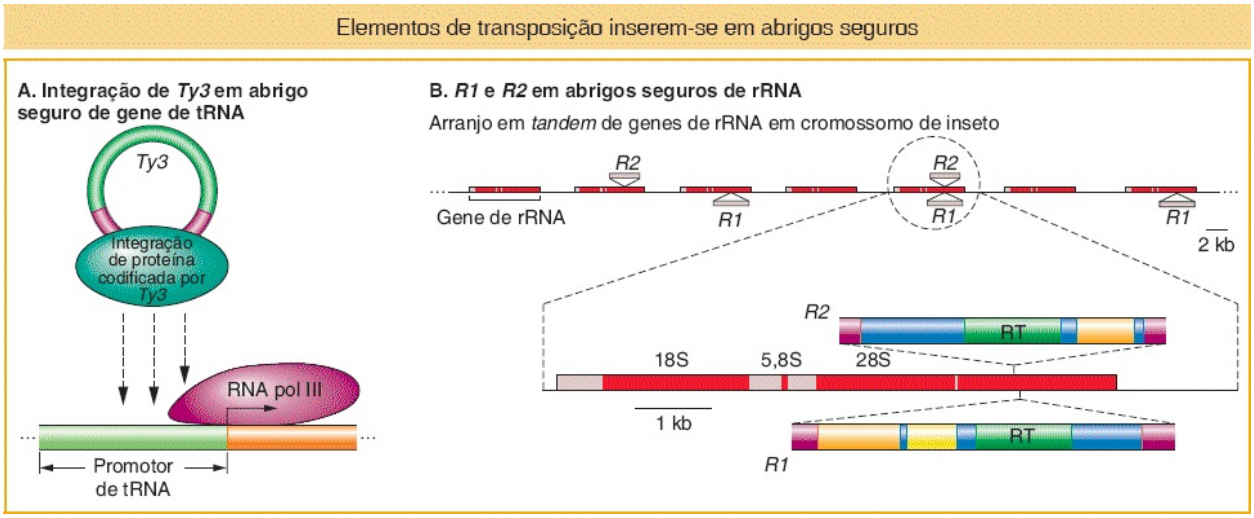


FIGURA 15.24 Alguns elementos de transposição são direcionados a abrigos seguros específicos. A. O retrotranspóson *Ty3* de levedura insere-se na região promotora de genes de RNA transportador. B. Os retrotranspósons não LTR (LINE) *R1* e *R2* de *Drosophila* inserem-se em genes que codificam RNA ribossômico que são observados em longos arranjos em *tandem* no cromossomo. Apenas os genes de transcriptase reversa (RT) de *R1* e *R2* são destacados.

Claramente, o sério risco associado a esse tipo de terapia gênica poderia ser bem menor se os médicos pudessem controlar onde o vetor retroviral se integra no genoma humano. Já vimos que existem muitas semelhanças entre os retrotranspósons LTR e os retrovírus. Existe a esperança de que, ao compreendermos o direcionamento de *Ty* em leveduras, possamos aprender como

construir vetores retrovirais que se inseriram, com sua carga transgênica, em abrigos seguros no genoma humano.

> CONCEITO-CHAVE Um elemento de transposição bem-sucedido aumenta o número de cópias sem prejudicar seu hospedeiro. Um modo por meio do qual um elemento aumenta com segurança o número de cópias é com o direcionamento de novas inserções para abrigos seguros, regiões do genoma nas quais existem poucos genes.

15.5 Regulação epigenética de elementos de transposição pelo hospedeiro

A análise genética é uma ferramenta muito poderosa utilizada para dissecar processos biológicos complexos por meio do isolamento de mutantes e, finalmente, de genes mutantes (p. ex., ver Capítulo 13 sobre o desenvolvimento). Muitos laboratórios em todo o mundo estão utilizando análises genéticas para identificar os genes hospedeiros responsáveis pela repressão de movimento dos elementos de transposição e, desse modo, manter a estabilidade do genoma.

A repressão de elementos de transposição foi investigada pela primeira vez no laboratório de Ron Plasterk no final da década de 1990, com a utilização do organismo-modelo *C. elegans* (um nematódeo; ver Organismo-modelo, no Capítulo 13). Essa história tem início com a observação de uma diferença surpreendente entre a mobilidade de um elemento de transposição denominado *Tc*1 em dois tipos celulares diferentes do organismo-modelo. O *Tc*1 é um transpóson de DNA que, assim como o elemento *Ac* do milho, pode levar a um fenótipo mutante instável quando é excisado de um gene com um fenótipo visível (ver Figura 15.4). Existem 32 elementos *Tc*1 no genoma sequenciado da linhagem de laboratório comum denominada Bristol. Significativamente, *Tc*1 é transposto em células somáticas, mas não nas células da linhagem germinativa. Tal observação sugeriu a Plasterk que a transposição é reprimida na linhagem

germinativa pelo hospedeiro. Evidentemente, a repressão na linhagem germinativa resulta do silenciamento dos genes de transposase de todas as 32 cópias do *Tc*1 nas células da linhagem germinativa. Você consegue propor uma explicação sobre qual o sentido biológico de um hospedeiro reprimir a transposição na linhagem germinativa e não na somática?

Plasterk e seus colaboradores decidiram identificar os genes de *C. elegans* responsáveis pelo silenciamento do gene da transposase. Eles iniciaram com uma linhagem de *C. elegans* que apresentava *Tc*1 inserido no gene *unc-22* (designado *unc-22/Tc*1; Figura 15.25). Esse era o mesmo gene que estava silenciado no experimento de Fire e Mello que os levou a compartilharem o Prêmio Nobel (ver Capítulo 8). Enquanto *C. elegans* do tipo selvagem desliza suavemente sobre a superfície do ágar em uma placa de Petri (conforme ilustrado pelas *setas horizontais* na Figura 15.25), vermes com o gene *unc-22/Tc*1 mutante apresentam um movimento de contração (conforme ilustrado pelas *setas verticais* na Figura 15.25), que pode ser facilmente observado com um microscópio. Tendo em vista que *Tc*1 normalmente não consegue realizar transposição na linhagem germinativa, ele permanece inserido no gene *unc-22* e continua a interromper a sua função. Portanto, a linhagem com o gene mutante *unc-22/Tc*1 deve expressar um fenótipo de contração de geração à geração. Entretanto, Plasterk e colaboradores raciocinaram que as mutações que inativavam os genes de *C. elegans* necessários para a repressão possibilitariam a *Tc*1 excisar-se do alelo *unc-22/Tc*1 na linhagem germinativa e reverter o fenótipo de contração para o tipo selvagem (*unc-22*). Com essa finalidade, eles expuseram a linhagem *unc-22/Tc*1 mutante a uma substância química que aumentava muito a frequência de mutação (denominada mutágeno; ver Capítulo 16) e examinaram a sua progênie ao microscópio, procurando por vermes raros que haviam deixado de apresentar a contração.

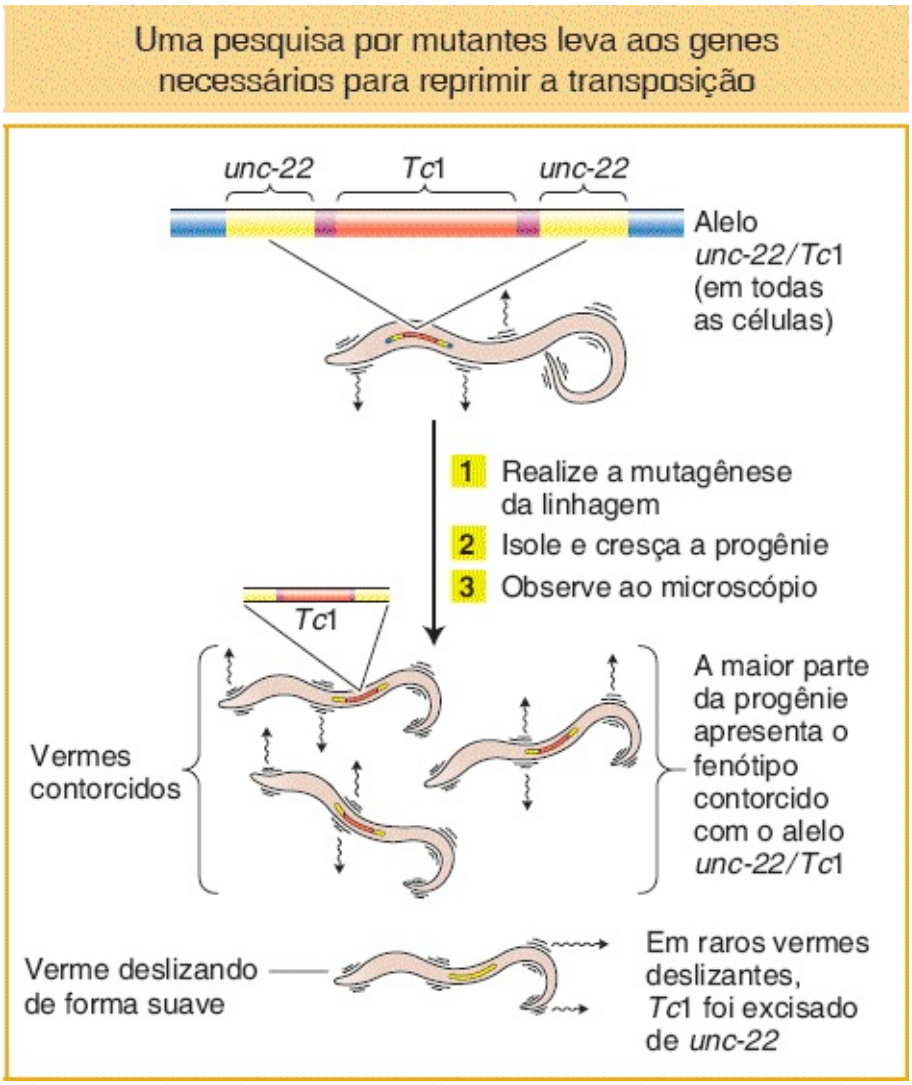


FIGURA 15.25 Desenho experimental utilizado para identificar os genes necessários para reprimir a transposição. Investigadores procuram por mutantes que tenham obtido novamente o movimento normal, tendo em vista que mutações nestes indivíduos teriam incapacitado o mecanismo de repressão que evita a transposição do elemento *Tc*1 do gene *unc-22*.

Essa triagem genética e outras subsequentes identificaram mais de 25 genes de *C. elegans* que, quando mutados, possibilitavam que o hospedeiro excisasse o *Tc*1 na linhagem germinativa. Significativamente, muitos dos produtos desses genes são componentes integrantes da via de silenciamento do RNAi, incluindo proteínas observadas em Dicer e RISC (ver Capítulos 8 e 12). Lembre, do Capítulo 8, que Dicer se liga a dsRNA longos e os cliva em pequenos fragmentos de dsRNA. Esses fragmentos em seguida se desenrolam de modo que um

filamento, o siRNA, possa direcionar RISC para cortar mRNA complementares (ver Figura 8.24).

Iniciando com essa elegante triagem genética, muitos anos de experimentação levaram ao modelo a seguir de repressão de elementos de transposição na linhagem germinativa de *C. elegans*. Com 32 elementos *Tc*1 dispersos por todo o genoma de *C. elegans*, alguns elementos próximos de genes são transcritos com o gene próximo por meio de transcrição "contínua de leitura" (ver Capítulo 8). Tendo em vista que as extremidades de *Tc*1 são repetições terminais invertidas de 54 pb, o RNA de *Tc*1 forma espontaneamente dsRNA (Figura 15.26). Assim como todos os dsRNA produzidos na maior parte dos eucariotos, esse RNA é reconhecido por Dicer e finalmente o siRNA é produzido, que direciona RISC para cortar os transcritos complementares de *Tc*1. Tendo em vista que todo RNA de *Tc*1 é cortado de modo eficiente na linhagem germinativa, o gene da transposase codificado pelo elemento é silenciado. Sem a transposase, o elemento não consegue excisar. Foi formulada a hipótese de que *Tc*1 consegue realizar transposição em células somáticas em virtude de o RNAi não ser tão eficiente e alguma transposase poder ser produzida.

Ao longo da última década, diversos laboratórios que trabalham com plantas e animais descobriram que as mutações que rompem a via do RNAi com frequência levam à ativação de elementos de transposição em seus respectivos genomas. Em virtude da abundância de elementos de transposição nos genomas eucarióticos e sua associação ancestral, sugeriu-se que a função natural da via do RNAi é manter a estabilidade do genoma ao reprimir o movimento dos elementos de transposição.

> CONCEITO-CHAVE Hospedeiros eucarióticos utilizam RNAi para reprimir a expressão de elementos de transposição ativos em seus genomas. Desse modo, um único elemento que se insere próximo de um gene pode ser transcrito para produzir dsRNA que acionará o silenciamento de todas as cópias do elemento no genoma.

Vigilância do genoma em animais e bactérias

A via de silenciamento do RNAi é semelhante a um radar, no sentido de que o hospedeiro é capaz de detectar novas inserções de transpósons no genoma se elas gerarem RNA antissenso. Em seguida, o hospedeiro responde por meio da produção de siRNA que tem por alvo o mRNA da transposase, silenciando o gene e prevenindo o movimento de todos os membros da família de elementos de transposição (TE). Recentemente, foram descritos dois outros tipos de radar do genoma (também denominado **vigilância do genoma**) que utilizam diferentes classes de pequenos RNA não codificadores com alvo em ácidos nucleicos "invasores", incluindo transpósons e vírus (ver Capítulo 5). Embora ainda não completamente compreendidos, esses mecanismos são apresentados aqui em virtude de ilustrarem como diferentes soluções são desenvolvidas para solucionar problemas biológicos semelhantes.

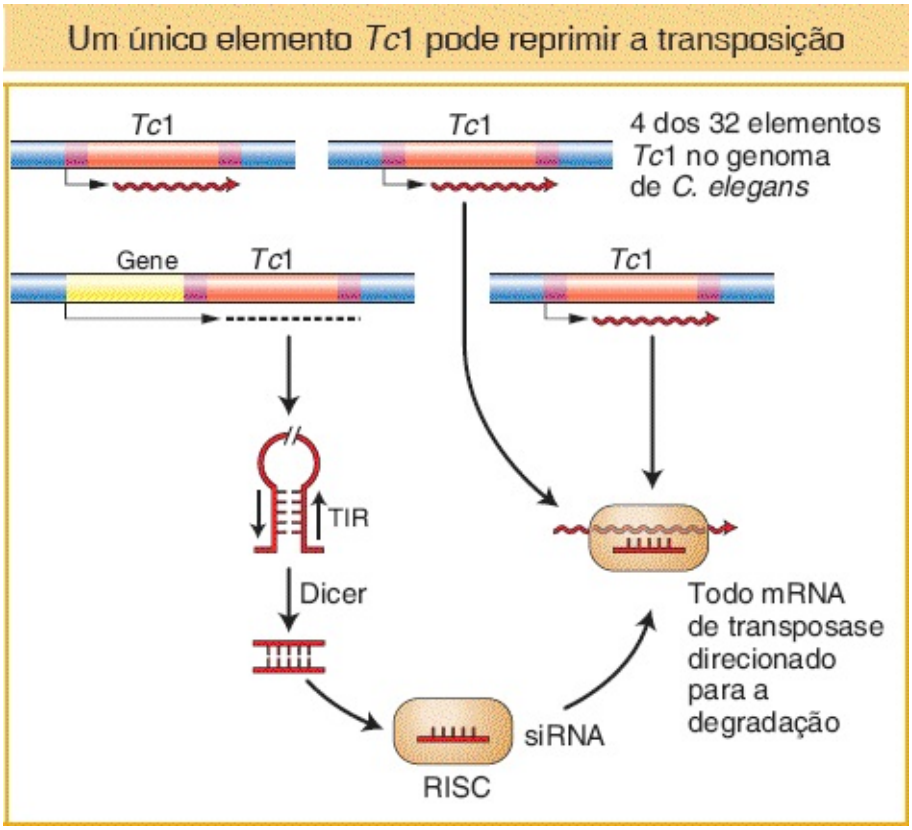


FIGURA 15.26 A produção de dsRNA de apenas um único elemento *Tc*1 é suficiente para silenciar todos os genes da transposase de *Tc*1 e, assim, reprimir a transposição na linhagem germinativa. O siRNA derivado do

dsRNA do *Tc*1 é ligado ao RISC que direciona todo o RNA complementar para degradação.

piRNA em animais. Nas linhagens germinativas de diversas espécies de animais, incluindo *Drosophila* e mamíferos, os transpósons ativos são reprimidos por meio da ação de **piRNA** (abreviação para RNA de interação *piwi*). Assim como os siRNA, os piRNA são RNA unifilamentares curtos (26 a 30 nt em mamíferos) que interagem com um complexo proteico (nesse caso, um que contém a proteína Piwi-Argonauta, da qual a sua denominação deriva). Uma vez associados, os piRNA orientam Piwi-Argonauta a degradar os mRNA complementares (Figura 15.27). Contrariamente aos siRNA, os piRNA não têm origem na via do RNA bifilamentar demonstrada na Figura 15.26 em relação a *Tc*1. Em vez disso, e de modo razoavelmente ingênuo, os genomas de animais contêm diversos *loci* longos (com frequência > 100 kb), denominados **agrupamentos pi**, que atuam como armadilhas para apanhar transpósons ativos. Um agrupamento pi é composto por remanescentes de muitos transpósons diferentes que representam um registro histórico das inserções anteriores de transpósons ativos naquele *locus*.

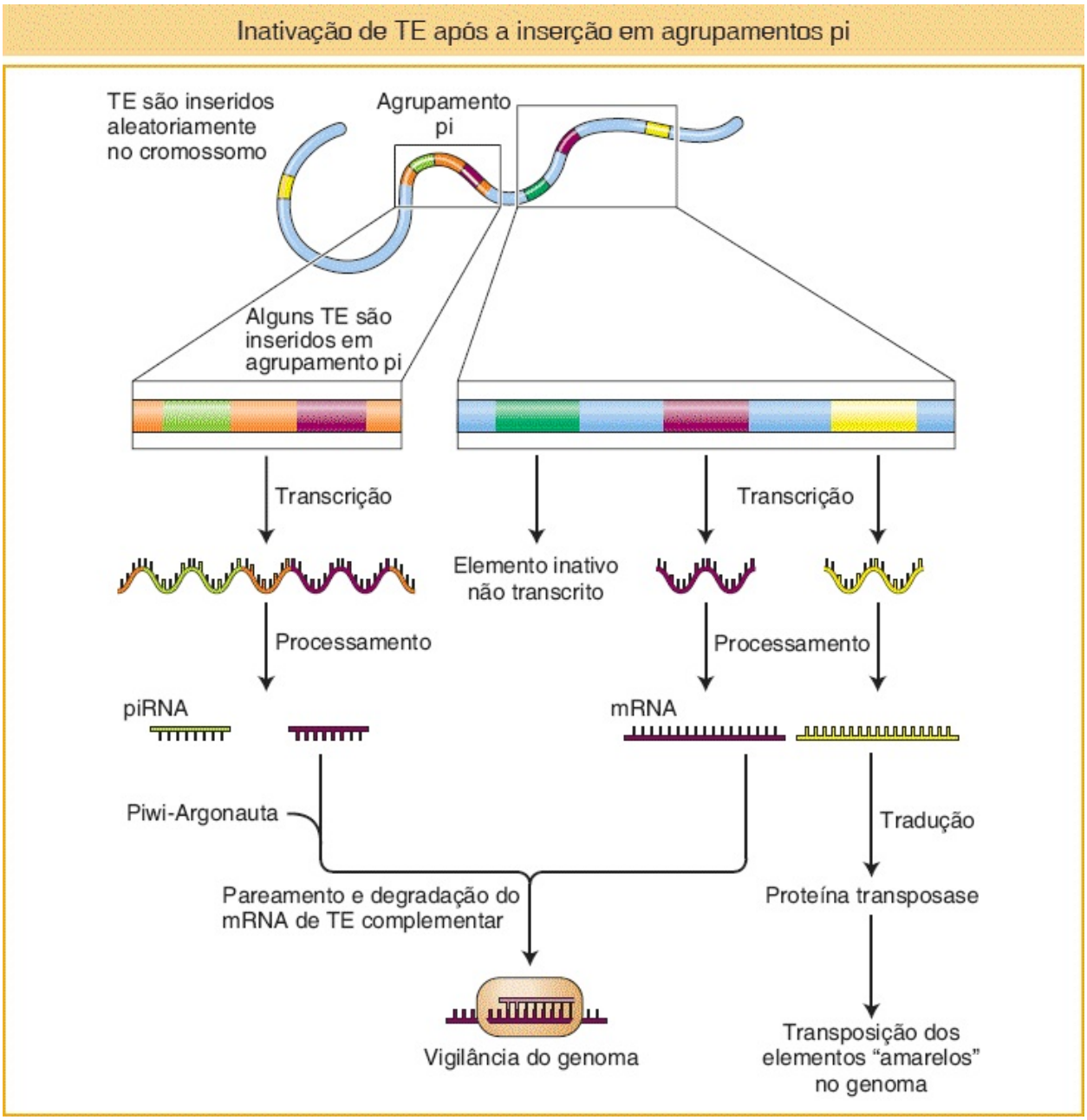


FIGURA 15.27 A inserção dos transpósons *verde* e *cor-de-rosa* em um agrupamento pi no genoma resulta na degradação de transcritos destes dois transpósons por meio das etapas demonstradas e descritas no texto. Contrariamente, o transpóson *amarelo* permanecerá ativo até que cópias sejam inseridas ao acaso em um agrupamento pi.

A primeira etapa na vigilância do genoma hospedeiro é a inserção de um transpóson em um de diversos *loci* de agrupamentos pi dispersos pelo genoma. A transcrição de agrupamentos pi produz RNA longos que podem incluir o RNA

antissenso do elemento recentemente inserido. Em seguida esses RNA longos são processados em piRNA finais que se associam com o Piwi-Argonauta e prosseguem para degradar mRNA derivados de transpósons transcritos a partir de qualquer local no genoma. Portanto, um TE ativo que se insere aleatoriamente no genoma é reconhecido pela vigilância do genoma apenas quando sua inserção ocorre em um agrupamento pi e ele se torna uma parte permanente do *locus*.

crRNA em bactérias. As moléculas de ácido nucleico normalmente invadem espécies bacterianas durante a infecção por bacteriófago (ver Capítulo 5) quando genomas de DNA viral são injetados em bactérias (ver Figura 5.22). Em uma via antiviral que ainda está sendo elucidada, fragmentos do genoma do vírus invasor são capturados por *loci* bacterianos denominados **CRISPR** (repetições palindrômicas curtas interespaçadas regularmente agrupadas) (Figura 15.28), nos quais são transcritos em RNA longos que são processados em **crRNA** curtos. De modo muito semelhante aos siRNA e aos piRNA, os crRNA interagem com e orientam os complexos proteicos bacterianos a degradar os RNA complementares do genoma viral invasor.

Uma característica compartilhada dos agrupamentos pi e dos *loci* CRISPR é que novas inserções de transpósons ou fragmentos de DNA viral, respectivamente, resultam em uma alteração genética permanente nesses *loci* que é herdada por suas progênies.

> CONCEITO-CHAVE Assim como siRNA, piRNA em animais e crRNA em bactérias interagem com complexos proteicos e os orientam a degradar sequências complementares em transpósons e vírus, respectivamente. Esses pequenos RNA não codificadores têm a sua origem em RNA longos transcritos a partir de *loci* que capturam fragmentos de DNA invasor.

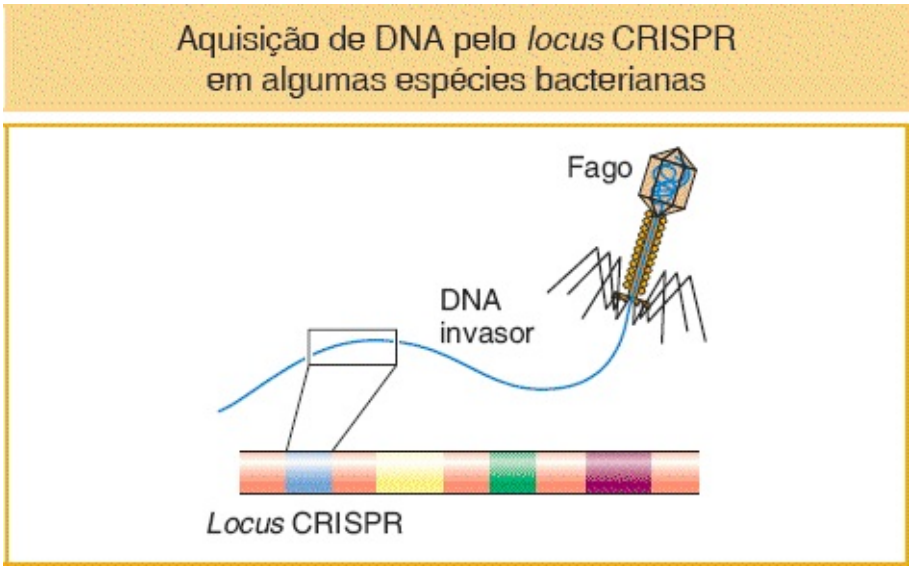


FIGURA 15.28 Aquisição de DNA pelo *locus* CRISPR em algumas espécies bacterianas. Parte do DNA do genoma de um fago invasor (demonstrado em *amarelo*) é incorporada ao *locus* CRISPR por meio de um mecanismo desconhecido.

O QUE OS GENETICISTAS FAZEM ATUALMENTE

De modo muito semelhante a aviões que escapam de radares ao voar próximo do solo, alguns transpósons desenvolveram mecanismos que os possibilitam escapar da via de silenciamento do RNAi. Esses transpósons conseguem alcançar números de cópias muito altos. As evidências em relação a esses mecanismos podem ser observadas nos genomas de todas as plantas e de todos os animais caracterizados que contêm famílias de transpósons (tais como *Alu* em seres humanos) que apresentam milhares de membros.

Como alguns transpósons evitam a detecção pela via de silenciamento do RNAi? A resposta curta é que, na maior parte dos casos, não sabemos. Para compreender como um transpóson evita ser detectado, é necessário estudar ativamente os elementos em transposição. Até o momento, os cientistas detectaram muito poucas famílias de transpósons com altos números de cópias que ainda estão ativamente em transposição. Um dos elementos mais bem-caracterizados desse pequeno grupo é um tipo especial de transpóson de DNA não autônomo denominado **elemento transponível miniatura com repetições**

invertidas (**MITE;** ***miniature inverted repeat transposable element***). Assim como outros elementos não autônomos, os MITE podem formar-se por deleção do gene da transposase a partir de um elemento autônomo. Entretanto, contrariamente à maior parte dos elementos não autônomos, os MITE podem alcançar números de cópias muito altos, particularmente nos genomas de alguns gramíneas (ver Figura 15.23). Alguns MITE em gramíneas foram amplificados até milhares de cópias.

O único MITE ativamente em transposição isolado até o momento é o elemento *mPing* do arroz, que é formado a partir do elemento *Ping* autônomo por deleção de todo o gene da transposase (Figura 15.29). Esse elemento foi descoberto no laboratório de Susan Wessler por Ning Jiang. Outro membro do laboratório de Wessler, Ken Naito, documentou que, em algumas linhagens de arroz, existem apenas 3 a 7 cópias de *Ping* e mais de 1.000 cópias de *mPing*. Notavelmente, o número de cópias de *mPing* nessas linhagens está aumentando em quase 40 novas inserções a cada planta por geração.

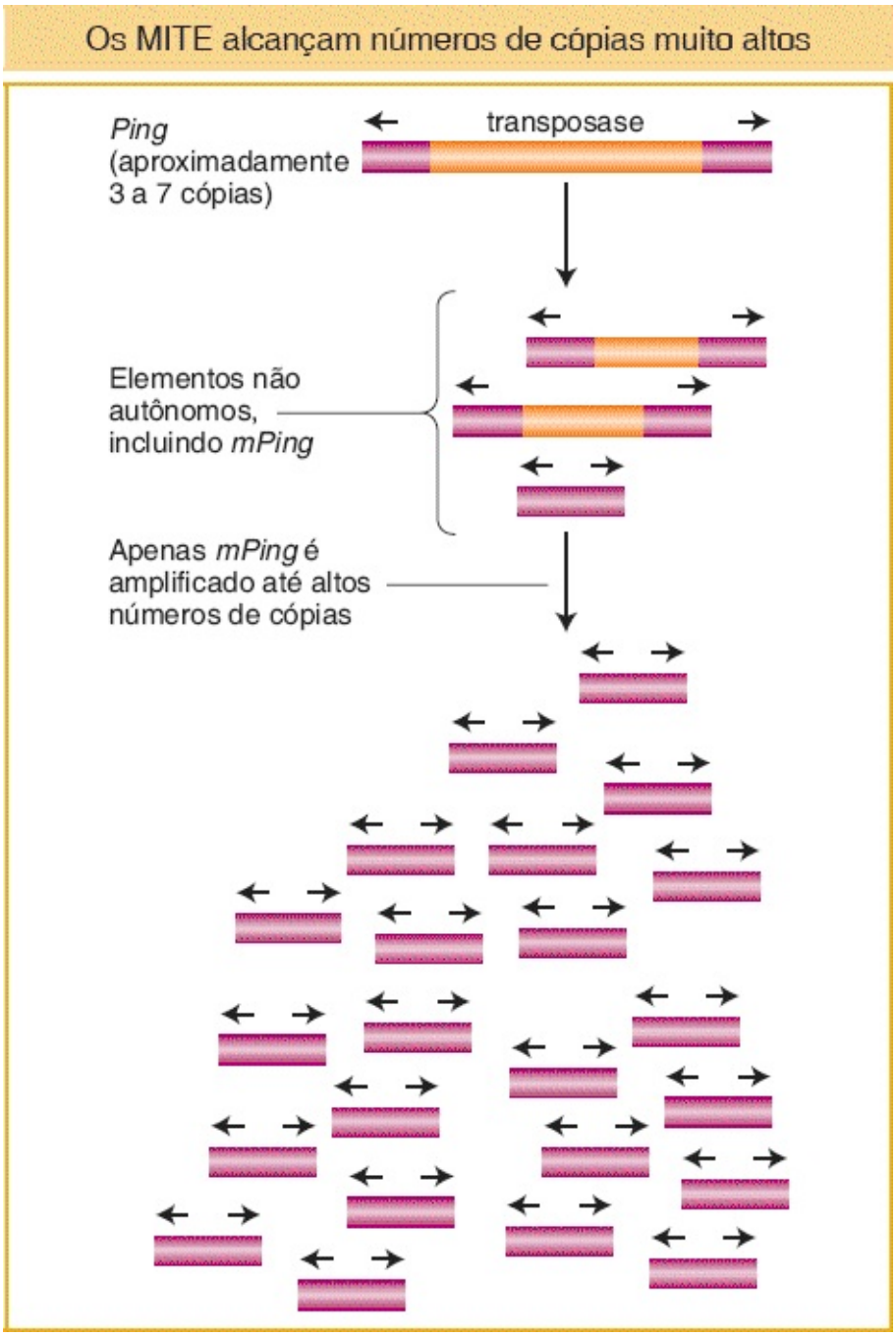


FIGURA 15.29 MITE são transpósons de DNA não autônomos que podem alcançar números de cópias muito altos, tendo em vista que não codificam a transposase necessária para a sua transcrição. O MITE *mPing* ativo é o único derivado por deleção do elemento *Ping* autônomo que alcançou um alto número de cópias em determinadas linhagens de arroz.

Duas questões a respeito do rápido aumento no número de cópias de *mPing* surgem imediatamente no pensamento. Primeiramente, como uma linhagem de

arroz sobrevive a uma explosão de elementos de transposição dessa magnitude? Para abordar essa questão, o laboratório de Wessler utilizou a tecnologia de sequenciamento de nova geração (ver Capítulo 14) para determinar os sítios de inserção de mais de 1.700 elementos *mPing* no genoma do arroz. Surpreendentemente, eles observaram que o elemento evitava inserir-se em éxons, minimizando, assim, o impacto da inserção sobre a expressão gênica do arroz. O mecanismo de base dessa preferência está sendo investigado atualmente.

A segunda questão é: por que o arroz hospedeiro aparentemente falha em reprimir a transposição do *mPing*? Embora essa questão também seja uma área ativa de pesquisas atuais, uma hipótese razoável é que o *mPing* consiga voar abaixo do radar do RNAi dos hospedeiros, tendo em vista que ele não contém qualquer parte do gene da transposase que está localizado no elemento *mPing* (Figura 15.29). Portanto, a transcrição de leitura nos elementos *mPing* inseridos por todo o genoma do arroz produzirá muitos lotes de dsRNA e siRNA. Entretanto, tendo em vista que os siRNA derivados de *mPing* não compartilham sequência com a fonte de transposase, os siRNA produzidos a partir de *mPing* não induzem os mecanismos de silenciamento direcionados à transposase. Em vez disso, o gene da transposase permanecerá ativo e continuará a catalisar a movimentação de *mPing*. De acordo com essa hipótese, a transposição de *mPing* será reprimida apenas quando uma inserção de *Ping* muito mais rara gerar dsRNA que acione o silenciamento de seu gene da transposase.

> CONCEITO-CHAVE MITE são transpósons de DNA não autônomos que conseguem alcançar altos números de cópias. Embora os MITE possam utilizar a transposase de elementos autônomos, eles provavelmente escapam à repressão do hospedeiro em virtude de sua amplificação não levar ao silenciamento do gene da transposase.

RESUMO

Os elementos de transposição foram descobertos em milho por Barbara

McClintock como responsáveis por diversas mutações instáveis. Um exemplo de um elemento não autônomo é *Ds*, cuja transposição requer a presença do elemento autônomo *Ac* no genoma.

As sequências de inserção de elementos bacterianos foram os primeiros elementos de transposição molecularmente isolados. Existem muitos tipos diferentes de elementos IS em linhagens de *E. coli* e eles normalmente estão presentes em diversas cópias. Os transpósons compostos contêm elementos IS flanqueando um ou mais genes, tais como os genes que conferem a resistência a antibióticos. Os transpósons com genes de resistência conseguem se inserir em plasmídios e em seguida são transferidos por meio de conjugação para bactérias não resistentes.

Existem dois grupos principais de elementos de transposição em eucariotos: elementos da classe 1 (retrotranspósons) e elementos da classe 2 (transpósons de DNA). O elemento *P* foi o primeiro transpóson de DNA eucariótico da classe 2 a ser molecularmente isolado. Ele foi isolado a partir de mutações instáveis em *Drosophila*, que foram induzidas por disgenesia híbrida. Os elementos *P* foram desenvolvidos em vetores para a introdução de DNA exógeno nas células germinativas de *Drosophila*.

Ac, *Ds* e *P* são exemplos de transpósons de DNA, assim denominados em virtude de o intermediário da transposição ser o próprio elemento de DNA. Elementos autônomos tais como *Ac* codificam uma transposase que se liga às extremidades de elementos autônomos e não autônomos, e catalisa a excisão do elemento do sítio doador e a reinserção em um novo sítio-alvo em algum outro local no genoma.

Os retrotranspósons foram molecularmente isolados pela primeira vez a partir de mutantes de leveduras e a sua semelhança com os retrovírus ficou logo evidente. Os retrotranspósons são elementos da classe 1, assim como o são todos os elementos de transposição que utilizam RNA como seu intermediário de transposição.

Os elementos de transposição ativos isolados a partir de organismos-modelo tais como levedura, *Drosophila*, *E. coli* e milho constituem uma fração muito pequena de todos os elementos de transposição no genoma. O sequenciamento do

DNA de genomas inteiros, incluindo o genoma humano, levou ao achado surpreendente de que quase metade do genoma humano é derivado de elementos de transposição. Apesar da apresentação de tantos elementos de transposição, os genomas eucarióticos são extremamente estáveis, assim como a transposição é relativamente rara em virtude de dois fatores. Primeiramente, a maior parte dos elementos de transposição em genomas eucarióticos não consegue mover-se, tendo em vista que mutações inativadoras evitam a produção da transposase e da transcriptase reversa normais. Em segundo lugar, a expressão da maioria dos elementos remanescentes é silenciada pela via do RNAi em plantas e *C. elegans* e da via do piRNA em animais. O silenciamento de elementos de transposição depende da capacidade do hospedeiro de detectar novas inserções no genoma e de gerar pequenos RNA não codificadores que orientam os complexos proteicos a degradarem RNA complementares codificados por transpósons. Alguns elementos com altos números de cópias, tais como os MITE, podem escapar do silenciamento em virtude de não acionarem o silenciamento da transposase que catalisa a sua transposição.

TERMOS-CHAVE

abrigo seguro
agrupamento pi
Alu
Ativador (*Ac*)
citotipo M
citotipo P
cointegrado
“cópia e colagem”
“corte e colagem”
crRNA
direcionamento
disgenesia híbrida
Dissociação (*Ds*)

duplicação de sítio-alvo
elemento autônomo
elemento curto intercalado (SINE)
elemento de sequência de inserção (IS)
elemento de transposição da classe 1 (retrotransposon)
elemento de transposição da classe 2 (transpóson de DNA)
elemento longo intercalado (LINE)
elemento não autônomo
elemento *P*
elemento tipo *copia*
elemento transponível miniatura com repetições invertidas (MITE)
elemento *Ty*
excisado
factor R
fenótipo instável
loci CRISPR
longa repetição terminal (LTR)
LTR solo
marcação de transpóson
paradoxo do valor *C*
piRNA
provírus
repetição invertida
retrotranspóson
retrotranspóson LTR
retrovírus
seleção negativa
sintenia
terapia gênica
transcriptase reversa
transposase
transposição

transposição conservativa
transposição replicativa
transpóson (Tn)
transpóson composto
transpóson de DNA
transpóson simples
transposto
valor *C*
vigilância do genoma

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Os elementos de transposição foram denominados “genes saltadores” em virtude de aparentemente pularem de uma posição para outra, deixando o *locus* antigo e aparecendo em um novo *locus*. À luz do que agora sabemos a respeito do mecanismo de transposição, quão apropriado é o termo “genes saltadores” em relação aos elementos de transposição bacterianos?

Solução

Em bactérias, a transposição ocorre por meio de dois modos diferentes. O modo conservativo resulta em genes saltadores verdadeiros, tendo em vista que, nesse caso, o elemento de transposição é excisado de sua posição original e é inserido em uma nova posição. O outro modo é o modo replicativo. Nessa via, um elemento de transposição move-se até uma nova localização por meio da replicação no DNA-alvo, deixando para trás uma cópia do elemento de transposição no sítio original. Ao operar por meio do modo replicativo, os elementos de transposição não são realmente genes saltadores, uma vez que uma cópia permanece no sítio original.

Problema resolvido 2. Após a questão anterior, à luz do que agora sabemos a respeito do mecanismo de transposição, quão apropriado é o termo “genes saltadores” em relação à maioria dos elementos de transposição no genoma

humano e nos genomas da maior parte dos outros mamíferos?

Solução

A maioria dos elementos de transposição nos genomas caracterizados de mamíferos é de retrotranspósons. Em seres humanos, dois retrotranspósons (o LINE denominado *L1* e o SINE denominado *Alu*) são responsáveis por até um terço do nosso genoma inteiro. Assim como os elementos bacterianos, os retrotranspósons não são excisados do sítio original, de modo que eles não são realmente genes saltadores. Em vez disso, o elemento atua como um molde para a transcrição de RNA, que podem ser transcritos de modo reverso pela enzima transcriptase reversa em um cDNA bifilamentar. Cada cDNA possivelmente pode inserir-se em sítios-alvo por todo o genoma. Observe que embora ambos os elementos bacterianos e os retrotranspósons não deixem o sítio original, seus respectivos mecanismos de transposição são dramaticamente diferentes. Finalmente, embora os retrotranspósons LTR não sejam excisados, eles podem se tornar inserções muito mais curtas em virtude da produção de LTR solo por meio de recombinação.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na fotografia de abertura do capítulo dos grãos em uma espiga de milho, qual é a base genética do que segue (**Dica**: consulte a Figura 15.4 para algumas indicações):

a. Do grão totalmente pigmentado?

b. Dos grãos não pigmentados? Observe que eles podem ser originados por dois modos diferentes.

2. Na Figura 15.3 A, qual seria o fenótipo do grão se a linhagem fosse homozigota em relação a todos os marcadores dominantes no cromossomo 9?

3. Em relação à Figura 15.7, desenhe uma série de etapas que possam explicar

a origem desse grande plasmídio, que contém muitos elementos de transposição.

4. Desenhe uma figura em relação ao modo de transposição não demonstrado na Figura 15.8, a retrotransposição.
5. Na Figura 15.10, demonstre onde a transposase precisaria realizar o corte para gerar uma duplicação no sítio-alvo de 6 pb. Demonstre também a localização do corte para gerar uma duplicação no sítio-alvo de 4 pb.
6. Se o elemento de transposição na Figura 15.14 fosse um transpóson de DNA que apresentasse um íntron no seu gene da transposase, o íntron seria removido durante a transposição? Justifique sua resposta.
7. Em relação à Figura 15.22, desenhe o pré-mRNA que é transcrito a partir desse gene e em seguida desenhe o seu mRNA.

PROBLEMAS BÁSICOS

8. Descreva a geração de plasmídios resistentes a múltiplos fármacos.
9. Descreva brevemente o experimento que demonstra que a transposição do elemento *Ty*1 em leveduras ocorre por meio de um RNA intermediário.
10. Explique como as propriedades dos elementos *P* em *Drosophila* tornam os experimentos de transferência gênica possíveis nesse organismo.
11. Embora os transpósons de DNA sejam abundantes nos genomas de eucariotos multicelulares, elementos da classe 1 normalmente compõem a maior fração de genomas muito grandes, tais como aqueles de seres humanos (aproximadamente 2.500 Mb), do milho (aproximadamente 2.500 Mb) e da cevada (aproximadamente 5.000 Mb). Tendo em vista o que você sabe a respeito dos elementos das classes 1 e 2, o que dizer sobre os seus mecanismos de transposição distintos que seriam responsáveis por essa diferença consistente na abundância?
12. Conforme você observou na Figura 15.22, os genes de eucariotos multicelulares com frequência contêm muitos elementos de transposição. Por que a maior parte desses elementos não afeta a expressão do gene?

13. O que são abrigos seguros? Existem locais nos genomas bacterianos muito mais compactos que podem ser um abrigo seguro para os elementos de inserção?

14. Os Prêmios Nobel normalmente são concedidos muitos anos após a descoberta real. Por exemplo, James Watson, Francis Crick e Maurice Wilkens receberam o Prêmio Nobel de Fisiologia ou Medicina em 1962, quase uma década após a sua descoberta da estrutura em dupla-hélice do DNA. Entretanto, Barbara McClintock recebeu o prêmio em 1983, quase quatro décadas após a sua descoberta dos elementos de transposição em milho. Por que você acredita que tenha demorado tanto para que a significância da sua descoberta fosse reconhecida dessa maneira?

15. A proteína transposase consegue:

a. Ligar-se ao DNA.

b. Catalisar a excisão de um elemento de transposição de um sítio doador.

c. Catalisar a inserção de um elemento de transposição em um sítio-alvo.

d. Todas as alternativas anteriores.

16. Quais dos seguintes são abrigos seguros para as inserções de elementos de transposição?

a. Íntrons.

b. Éxons.

c. Outros elementos de transposição.

d. As alternativas **a** e **c** estão corretas.

17. Por que os retrotranspósons não conseguem se mover de uma célula para outra, assim como os retrovírus?

a. Porque eles não codificam a proteína Env.

b. Porque eles são elementos não autônomos.

c. Porque eles necessitam da transcriptase reversa.

d. As alternativas **a** e **b** são verdadeiras.

18. Contrariamente aos retrotranspósons, os transpósons de DNA:

a. Apresentam repetições terminais invertidas.

b. Geram uma duplicação do sítio-alvo após a inserção.

c. São transpostos por meio de um RNA intermediário.

d. Não são observados em procariotos.

19. A principal diferença entre os retrotranspósons e os retrovírus é:

a. Os retrotranspósons codificam transcriptase reversa.

b. Os retrovírus se movem de um sítio no genoma até outro.

c. Os retrovírus codificam o gene *env*, que possibilita que eles se movimentem de uma célula para outra.

d. Nenhuma das alternativas anteriores está correta.

20. Qual das seguintes é verdadeira em relação à transcriptase reversa?

a. Ela é necessária para a movimentação dos transpósons de DNA.

b. Ela catalisa a síntese do DNA a partir do RNA.

c. Ela é necessária para a transposição de retrotranspósons.

d. As alternativas **b** e **c** estão corretas.

21. Qual elemento de transposição é utilizado para introduzir DNA exógeno na mosca-das-frutas *Drosophila melanogaster*?

a. Elemento *Ac*.

b. Elemento *P*.

c. Elemento *Alu*.

d. Transpósons compostos.

22. Qual é o principal motivo pelo qual o genoma de milho é muito maior que o genoma de arroz?

a. O milho apresenta mais genes do que o arroz.

b. O arroz apresenta mais genes do que o milho.

c. O milho apresenta mais transpósons de DNA do que o arroz.

d. O milho apresenta mais retrotranspósons do que o arroz.

23. Por que os elementos de transposição são observados com muito mais frequência em íntrons do que em éxons?

a. Porque os elementos de transposição preferem se inserir em íntrons.

b. Porque os elementos de transposição preferem se inserir em éxons.

c. Porque os elementos de transposição se inserem tanto em éxons como em íntrons, mas a seleção remove as inserções em éxons.

d. Nenhuma das alternativas anteriores é verdadeira.

24. Aproximadamente que porcentagem do genoma humano é derivada de elementos de transposição?

a. 10%.

b. 25%.

c. 50%.

d. 75%.

25. Por que as plantas e os animais prosperam com tantos elementos de transposição em seus genomas?

a. A maior parte dos elementos de transposição é inativa em virtude de mutação.

b. Os elementos de transposição ativos são silenciados pelo hospedeiro.

c. A maior parte dos elementos de transposição está inserida em abrigos seguros.

d. Todas as alternativas anteriores são verdadeiras.

PROBLEMAS DESAFIADORES

26. A inserção de elementos de transposição em genes pode alterar o padrão de expressão normal. Nas situações a seguir, descreva as possíveis consequências sobre a expressão gênica.

a. Um LINE inserido em um acentuador de um gene humano.

b. Um elemento de transposição contém um sítio de ligação para um repressor de transcrição e insere-se na posição adjacente a um promotor.

c. Um elemento *Alu* inserido no sítio de corte 3′ (AG) de um íntron em um gene humano.

d. Um elemento *Ds* que foi inserido no éxon de um gene de milho é excisado de modo imperfeito e deixa três pares de bases para trás no éxon.

e. Outra excisão por aquele mesmo elemento *Ds* deixa dois pares de bases para trás no éxon.

f. Um elemento *Ds* que foi inserido na parte intermediária de um íntron é excisado de modo imperfeito e deixa para trás cinco pares de bases no

íntron.

27. Antes da integração de um transpóson, sua transposase realiza um corte desencontrado no DNA-alvo do hospedeiro. Se o corte desencontrado está nos sítios das setas a seguir, desenhe qual será a sequência do DNA hospedeiro após o transpóson ter sido inserido. Represente o transpóson como um retângulo.

```
          ↓
AATTTGGCCTAGTACTAATTGGTTGG
TTAAACCGGATCATGATTAACCAACC
                  ↑
```

28. Em *Drosophila*, M. Green encontrou um alelo *singed* (*sn*) com algumas características incomuns. As fêmeas homozigotas em relação a esse alelo ligado ao X apresentam cerdas *singed*, mas apresentam diversas áreas de cerdas sn^+ (tipo selvagem) na cabeça, no tórax e no abdome. Quando essas moscas são cruzadas com machos *sn*, algumas fêmeas produzem apenas progênie *singed*, mas outras produzem tanto progênies *singed* quanto do tipo selvagem em proporções variáveis. Explique esses resultados.

29. Considere duas plantas do milho:

a. Genótipo C/c^m; Ac/Ac^+, no qual c^m é um alelo instável causado por uma inserção de *Ds*.

b. Genótipo C/c^m, no qual c^m é um alelo instável causado por uma inserção de *Ac*.

Quais fenótipos seriam produzidos e em quais proporções quando (1) cada planta fosse cruzada com um mutante com substituição de par de bases *c/c* e (2) a planta na parte *a* fosse cruzada com a planta na parte *b*? Presuma que *Ac* e *c* não estejam ligados, que a frequência de quebra cromossômica seja insignificante e que aquele mutante *c/C* seja Ac^+.

30. Você encontra sua amiga, uma cientista, na academia e ela começa a lhe contar sobre um gene de camundongo que ela está estudando no laboratório. O produto desse gene é uma enzima necessária para produzir a pelagem

marrom. O gene é denominado *FB* e a enzima é denominada enzima FB. Quando *FB* é mutante e não consegue produzir a enzima FB, a pelagem é branca. A cientista diz a você que ela isolou o gene de dois camundongos com pelagem marrom e que, surpreendentemente, observou que os dois genes diferem pela presença de um SINE de 250 pb (assim como o elemento Alu humano) no gene *FB* de um camundongo, mas não no gene do outro. Ela não compreende como essa diferença é possível, especialmente tendo em vista que ela determinou que ambos os camundongos produzem a enzima FB. Você pode lhe ajudar a formular uma hipótese que explique por que o camundongo ainda produz a enzima FB com um elemento de transposição em seu gene *FB*?

31. O genoma de levedura apresenta elementos de transposição da classe 1 (*Ty1*, *Ty2* e assim por diante), porém nenhum elemento da classe 2. Qual é um possível motivo pelo qual os elementos de DNA não obtiveram sucesso no genoma de levedura?

32. Além de *Tc*1, o genoma de *C. elegans* contém outras famílias de transpósons de DNA, tais como *Tc*2, *Tc*3, *Tc*4 e *Tc*5. Assim como *Tc*1, sua transposição é reprimida na linhagem germinativa, mas não em células somáticas. Preveja o comportamento desses elementos nas linhagens mutantes nas quais *Tc*1 deixou de ser reprimido em virtude de mutações na via do RNAi. Justifique a sua resposta.

33. Com base no mecanismo de silenciamento gênico, quais características dos elementos de transposição a via do RNAi explora para assegurar que os genes do próprio hospedeiro também não sejam silenciados?

34. Quais são as semelhanças e as diferenças entre os retrovírus e os retrotranspósons? Foi formulada a hipótese de que os retrovírus se desenvolveram a partir de retrotranspósons. Você concorda com esse modelo? Justifique a sua resposta.

35. Você isolou um elemento de transposição do genoma humano e determinou a sequência de seu DNA. Como você utilizaria essa sequência para determinar o número de cópias do elemento no genoma humano se você

acabou de receber um computador com uma conexão à Internet? (**Dica**: ver Capítulo 14.)

36. No ensejo da questão anterior, como você determinaria se outros primatas apresentavam um elemento semelhante nos seus genomas?

37. De todos os genes no genoma humano, aqueles com as inserções de *Alu* mais caracterizadas são aqueles que causam a hemofilia, incluindo diversas inserções nos genes do fator VIII e do fator IX. Com base nesse fato, seu colega formula a hipótese de que o elemento *Alu* prefere se inserir nesses genes. Você concorda? Qual outro motivo você fornece que também explica esses dados?

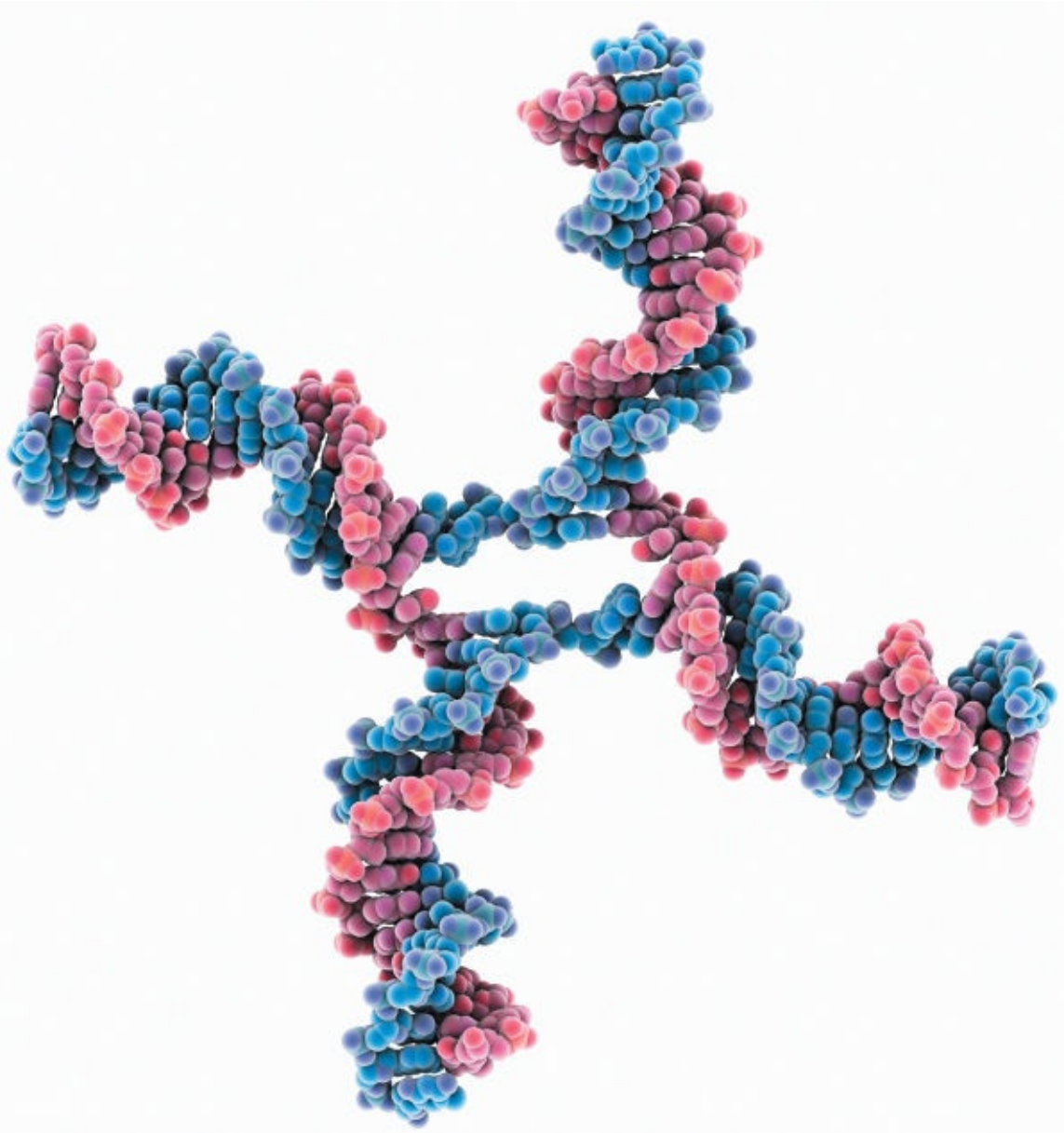
38. Se todos os membros de uma família de elementos de transposição podem ser silenciados pelo dsRNA sintetizado a partir de um único membro da família, como é possível que uma família de elementos (como *Tc*1) apresente 32 cópias no genoma do *C. elegans*, enquanto outra família (*Tc2*) apresenta menos de cinco cópias?

39. Como os *loci* CRISPR e de agrupamento pi são alterados ao longo do tempo?

CAPÍTULO

16

Mutação, Reparo e Recombinação



Um modelo computadorizado de dois cromossomos sofrendo um *crossover*. (Laguna Design/Science Photo Library/Science Source.)

TÓPICOS

16.1 Consequências fenotípicas das mutações no DNA
16.2 Base molecular das mutações espontâneas
16.3 Base molecular das mutações induzidas
16.4 Mecanismos biológicos de reparo
16.5 Câncer | Uma importante consequência fenotípica da mutação

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Explicar a base molecular das mutações
- Comparar e contrastar as origens e os efeitos das mutações espontâneas *versus* induzidas
- Descrever os diferentes mecanismos de reparo biológico
- Descrever as doenças genéticas humanas que são causadas por mutações nos mecanismos de reparo
- Discutir as diferenças entre as células cancerosas e normais
- Explicar por que os agentes mutagênicos podem causar alguns cânceres.

Uma paciente jovem desenvolve uma grande quantidade de pequenas lesões cutâneas pré-cancerosas semelhantes a sardas e é extremamente sensível à luz solar (Figura 16.1). É obtido um histórico familiar e a paciente é diagnosticada com uma doença autossômica recessiva denominada xeroderma pigmentoso. Durante toda a sua vida, ela estará propensa ao desenvolvimento de cânceres pigmentados de pele. Diversos genes diferentes podem ser mutados para gerar esse fenótipo de doença. Em uma pessoa sem a doença, cada um desses genes contribui para processos bioquímicos na célula que respondem a dano químico no DNA e o reparam antes que ele leve à formação de

novas mutações. Posteriormente neste capítulo, veremos como as mutações nos sistemas de reparo levam a doenças genéticas tais como o xeroderma pigmentoso.

Pessoas com essa doença são exemplos de *variantes* genéticas – indivíduos que demonstram diferenças fenotípicas em uma ou mais características em particular. Tendo em vista que a genética é o estudo das diferenças hereditárias, a análise genética não seria possível sem variantes. Nos capítulos precedentes, você viu muitas análises da herança de tais variantes; agora, consideraremos a sua origem. Como surgem as variantes genéticas?

Dois processos principais são responsáveis pela variação genética: *mutação* e *recombinação*. Vimos que a mutação é uma alteração na sequência de DNA de um gene. A mutação é especialmente significativa, tendo em vista que é a fonte primária de alteração evolutiva; novos alelos surgem em todos os organismos, alguns espontaneamente e outros resultantes de exposição à radiação ou às substâncias químicas no ambiente. Os novos alelos produzidos por mutação se tornam a matéria-prima para um segundo nível de variação, efetuado por recombinação. Conforme a sua denominação sugere, a recombinação é o desfecho de processos celulares que causam o agrupamento de alelos de diferentes genes em novas combinações (ver Capítulo 4). Para utilizar uma analogia, a mutação ocasionalmente produz uma nova carta nos jogos, mas é a recombinação que embaralha as cartas e as distribui de modos diferentes.

Câncer de pele no xeroderma pigmentoso

FIGURA 16.1 A doença hereditária recessiva xeroderma pigmentoso é causada por deficiência em uma de diversas proteínas que auxiliam a corrigir o DNA danificado. Essa deficiência enzimática leva à formação de cânceres de pele com a exposição da pele aos raios ultravioleta da luz solar. (KOKEL/BSIP/SuperStock.)

No ambiente celular, as moléculas de DNA não são absolutamente estáveis: cada par de bases em uma dupla-hélice de DNA apresenta uma determinada probabilidade de mutação. Conforme veremos, o termo *mutação* abrange uma ampla gama de tipos de alterações. Essas alterações variam desde a simples permuta de um par de bases por outro, até o desaparecimento de um cromossomo inteiro. No Capítulo 17, consideraremos as alterações mutacionais que afetam cromossomos inteiros ou grandes pedaços dos cromossomos. No presente capítulo, enfocamos os eventos mutacionais que ocorrem *em* genes individuais. Denominamos tais eventos mutações gênicas.

As células desenvolveram sistemas sofisticados para identificar e reparar DNA danificado, prevenindo, assim, a ocorrência da maior parte das mutações, mas não de todas. Podemos considerar que o DNA está sujeito a um cabo de guerra dinâmico entre os processos químicos que o danificam e o levam a novas mutações e os processos de reparo celular que o monitoram constantemente em relação a tais danos para corrigi-los. Entretanto, esse cabo de guerra não é direto.

Conforme já mencionado, as mutações fornecem a matéria-prima para a evolução e, portanto, a introdução de um baixo nível de mutações deve ser tolerada. Veremos que os sistemas de replicação do DNA e de reparo do DNA de fato podem introduzir mutações. Outros transformam mutações possivelmente devastadoras (tais como quebras bifilamentares) em mutações que podem afetar apenas um único produto gênico.

Veremos que a classe possivelmente mais séria de dano ao DNA, uma quebra bifilamentar, também é uma etapa intermediária no processo celular normal de recombinação por meio do *crossing over* meiótico. Portanto, podemos traçar paralelos entre a mutação e a recombinação em dois níveis. Primeiramente, conforme mencionado anteriormente, a mutação e a recombinação são as principais fontes de variação. Em segundo lugar, os mecanismos de reparo e recombinação do DNA apresentam algumas características em comum, incluindo a utilização de algumas das mesmas proteínas. Por esse motivo, exploraremos primeiramente os mecanismos de reparo do DNA e em seguida os compararemos ao mecanismo de recombinação do DNA.

16.1 Consequências fenotípicas das mutações no DNA

O termo **mutação de ponto** refere-se tipicamente à alteração de um único par de bases do DNA ou de um pequeno número de pares de bases adjacentes. Nesta seção, consideraremos os efeitos das referidas alterações no nível fenotípico. As mutações de ponto estão classificadas em termos moleculares na Figura 16.2, que demonstra os principais tipos de alterações do DNA e seus efeitos sobre a função proteica quando elas ocorrem na região codificadora de proteína de um gene.

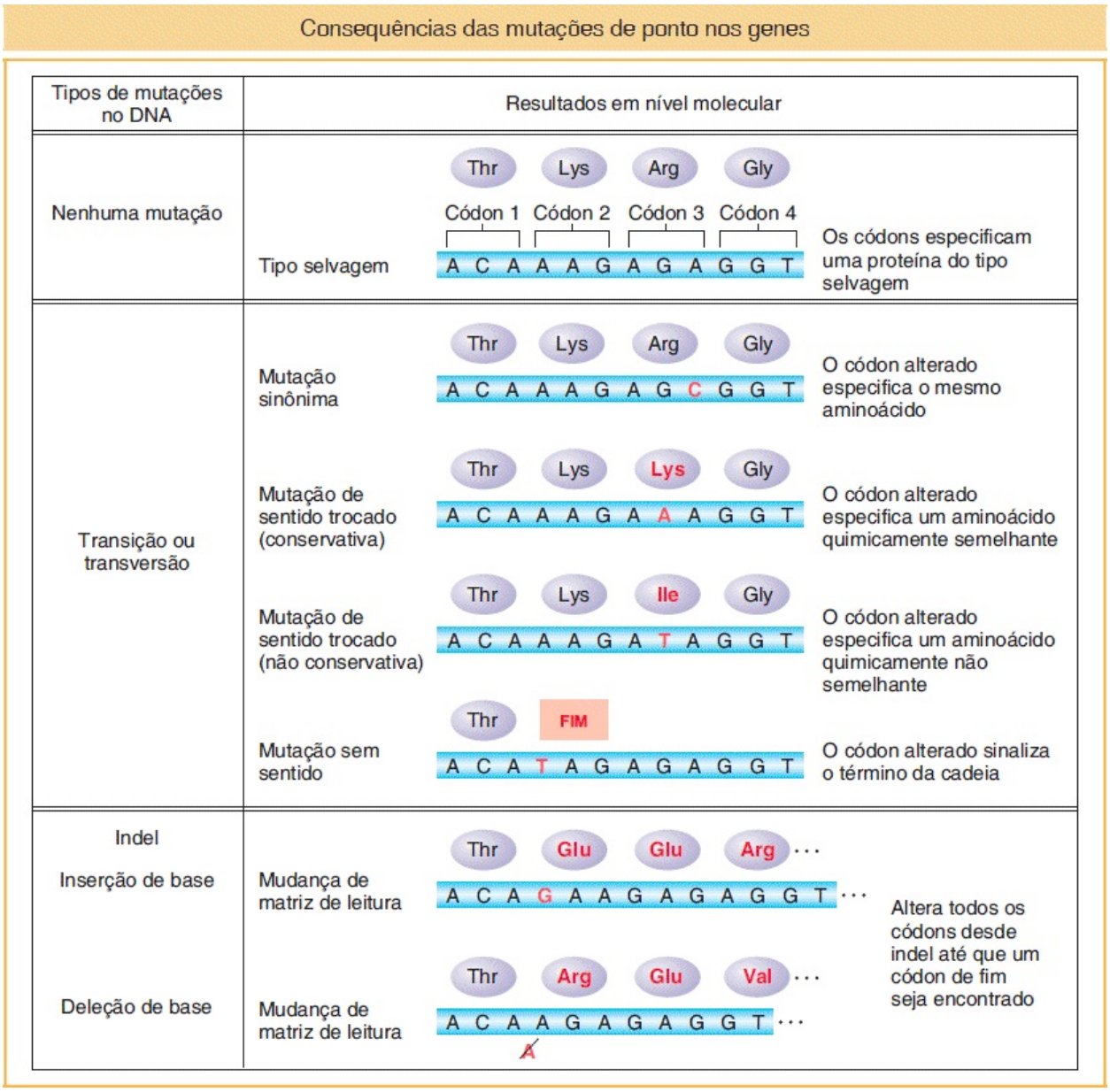


FIGURA 16.2 Mutações de ponto na região codificadora de um gene variam em seus efeitos sobre a função proteica. As proteínas com mutações sinônimas e de sentido trocado normalmente ainda são funcionais.

Tipos de mutação de ponto

Os dois tipos principais de mutação de ponto no DNA são as *substituições de bases* e as *inserções* ou *deleções de bases*. As substituições de bases são mutações nas quais um par de bases é substituído por outro, e podem ser divididas em dois subtipos: transições e transversões. Para descrever esses subtipos, consideremos como uma mutação altera a sequência de um filamento de

DNA (a alteração complementar ocorrerá no outro filamento). Uma **transição** é a substituição de uma base por outra base da mesma categoria química. Uma purina é substituída por uma purina (A por G ou G por A) ou uma pirimidina é substituída por uma pirimidina (C por T ou T por C). Uma **transversão** é o oposto – a substituição de uma base de uma categoria química por uma base da outra. Uma pirimidina é substituída por uma purina (C por A, C por G, T por A ou T por G) ou uma purina é substituída por uma pirimidina (A por C, A por T, G por C ou G por T). Ao descrever as mesmas alterações no nível do filamento duplo de DNA, devemos representar ambos os membros de um par de bases na mesma localização relativa. Portanto, um exemplo de uma transição é G · C → A · T; aquele de uma transversão é G · C → T · A. As mutações de inserção ou deleção são de fato inserções ou deleções de pares de *nucleotídios*; não obstante, a convenção é denominá-las inserções ou deleções de pares de *bases*. Coletivamente, elas são denominadas *mutações indel* (em referência a *in*serção-*del*eção). A mais simples dessas mutações é a adição ou a deleção de um único par de bases. As mutações por vezes surgem por meio da adição ou deleção simultânea de múltiplos pares de bases de uma só vez. Conforme veremos posteriormente neste capítulo, os mecanismos que produzem seletivamente determinados tipos de adições ou deleções de múltiplos pares de bases são a causa de determinadas doenças genéticas humanas.

Consequências moleculares das mutações de ponto em uma região codificadora

Quais são as consequências funcionais desses diferentes tipos de mutações de ponto? Primeiramente, consideramos o que ocorre quando surge uma mutação em uma parte de um gene codificadora de polipeptídio. Em relação às substituições de uma única base, existem diversos desfechos possíveis, mas todos são consequências diretas de dois aspectos do código genético: degeneração do código e existência de códons de término de tradução (ver Figura 16.2).

- **Mutações sinônimas.** A mutação altera um códon de um aminoácido por outro códon desse mesmo aminoácido. As mutações sinônimas em éxons também são

denominadas mutações *silenciosas*

- **Mutações de sentido trocado.** O códon de um aminoácido é alterado por um códon de outro aminoácido. As mutações de sentido trocado por vezes são denominadas mutações *não sinônimas*
- **Mutações sem sentido.** O códon de um aminoácido é alterado para um códon de término de tradução (fim).

As substituições sinônimas nunca alteram a sequência de aminoácidos da cadeia polipeptídica. A gravidade do efeito das mutações de sentido trocado e sem sentido no polipeptídio difere caso a caso. Por exemplo, uma mutação de sentido trocado pode substituir um aminoácido por um aminoácido quimicamente semelhante, sendo denominada substituição conservativa. Nesse caso, a alteração apresenta menor probabilidade de afetar gravemente a estrutura e a função da proteína. Alternativamente, um aminoácido pode ser substituído por outro quimicamente diferente, em uma **substituição não conservativa**. Esse tipo de alteração apresenta maior probabilidade de causar uma alteração grave na estrutura e na função da proteína. As mutações sem sentido levarão ao término antecipado da tradução. Portanto, elas apresentam um efeito considerável sobre a função proteica. Quanto mais próxima uma mutação sem sentido estiver da extremidade 3′ da matriz de leitura aberta (ORF), mais plausível será que a proteína resultante possa apresentar alguma atividade biológica. Entretanto, muitas mutações sem sentido produzem produtos proteicos completamente inativos.

Alterações de um único par de bases que inativam proteínas com frequência ocorrem em virtude de mutações no sítio de *corte*. Conforme observado na Figura 16.3, as referidas alterações podem modificar dramaticamente o transcrito de mRNA, levando a grandes inserções ou deleções, que podem ocorrer na matriz de leitura ou não.

Assim como as mutações sem sentido, as mutações indel podem apresentar consequências sobre a sequência de polipeptídios que se estendem até muito além do sítio da própria mutação (ver Figura 16.2). Relembre que a sequência de mRNA é “lida” pelo aparato de tradução na matriz de leitura, três bases (um

códon) por vez. A adição ou a deleção de um único par de bases do DNA altera a matriz de leitura do restante do processo de tradução, a partir do sítio da mutação do par de bases até o próximo códon de fim na nova matriz de leitura. Portanto, essas lesões são denominadas **mudança de matriz de leitura (***frameshift mutations***)**. Essas mutações fazem com que toda a sequência de aminoácidos *downstream* ao sítio mutante não apresente relação com a sequência de aminoácidos original. Portanto, as mudanças de matriz de leitura tipicamente resultam em perda completa da estrutura e da função proteica normal (Figura 16.4).

Consequências moleculares das mutações de ponto em uma região não codificadora

Agora nos voltaremos para as mutações que ocorrem nas sequências reguladoras e em outras não codificadoras. As partes de um gene que não codificam diretamente uma proteína contêm muitos sítios de ligação ao DNA cruciais para proteínas intercaladas entre as sequências que não são essenciais para a expressão ou atividade gênica. No nível do DNA, os sítios de ligação incluem aqueles aos quais a RNA polimerase e seus fatores associados se ligam, bem como os sítios aos quais proteínas específicas reguladoras da transcrição se ligam. No nível do RNA, sítios de ligação importantes adicionais incluem os sítios de ligação dos ribossomos de mRNA bacterianos, os sítios de corte 5′ e 3′ para a união de éxons em mRNA eucarióticos e sítios que regulam a tradução e localizam o mRNA em áreas e compartimentos particulares dentro da célula.

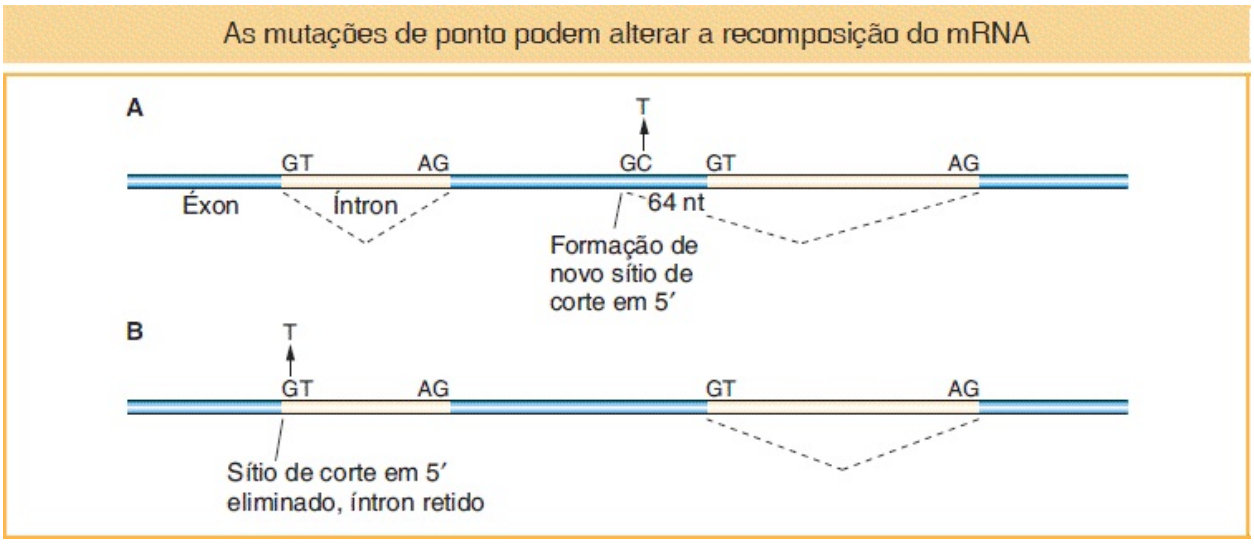


FIGURA 16.3 Dois exemplos demonstram as consequências das mutações de ponto em sítios de corte. **A.** Uma mutação por transição de C para T leva a um dinucleotídio GT no éxon, formando um novo sítio de corte em 5′. Como resultado, 64 nucleotídios na extremidade de um éxon são removidos. **B.** Uma mutação por transversão de G para T eliminaria o sítio de corte em 5′, de modo que o íntron seria retido no mRNA.

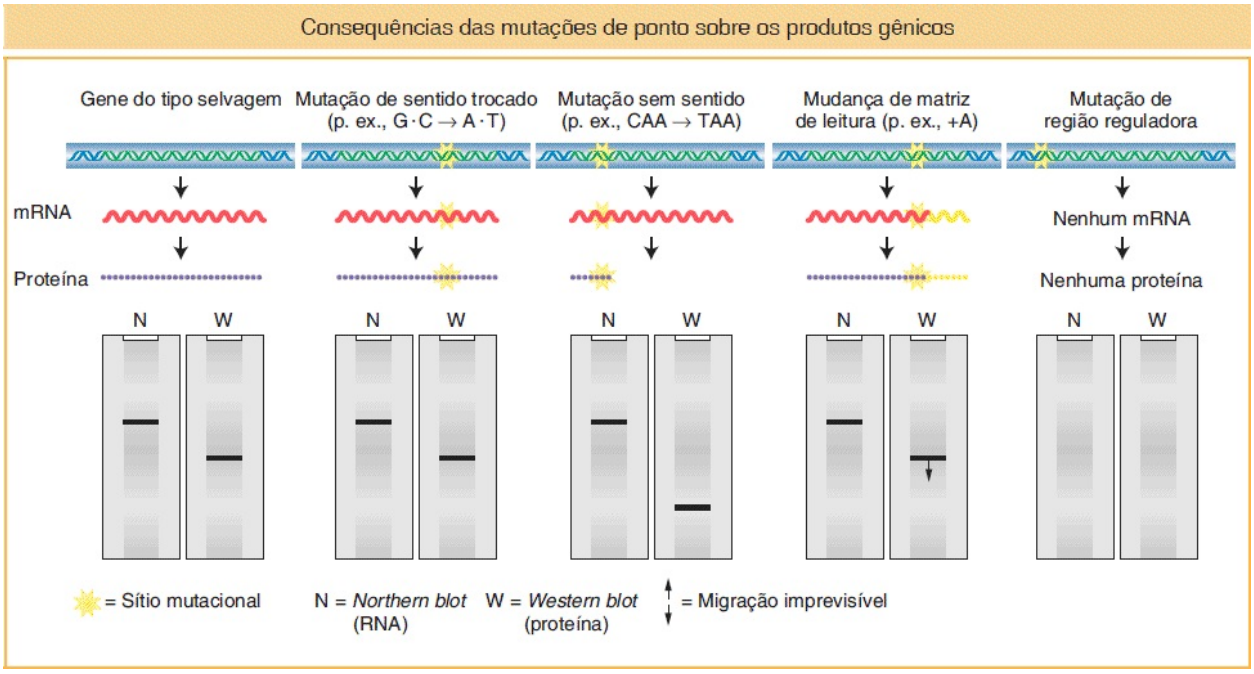


FIGURA 16.4 Mutações de ponto em regiões codificadoras podem alterar a estrutura da proteína, com ou sem alteração no tamanho do mRNA. As mutações de ponto em regiões reguladoras podem evitar a síntese de mRNA (e proteína).

As ramificações de mutações em partes de um gene além dos segmentos

codificadores de polipeptídios são mais difíceis de prever do que as mutações nos segmentos codificadores. Em geral, as consequências funcionais de qualquer mutação de ponto em tal região dependem do fato de a mutação abolir (ou criar) um sítio de ligação. As mutações que comprometem esses sítios apresentam o potencial de alterar o padrão de expressão de um gene ao alterar a quantidade do produto expresso em uma determinada ocasião ou em um determinado tecido ou ao alterar a resposta a determinados estímulos ambientais. As referidas mutações reguladoras irão alterar a quantidade de produto proteico produzida, mas *não* a estrutura da proteína. Alternativamente, algumas mutações em sítios de ligação podem bloquear completamente uma etapa necessária na expressão gênica normal (tal como a ligação da RNA polimerase ou dos fatores de recomposição) e, portanto, inativar totalmente o produto gênico ou bloquear a sua formação. A Figura 16.4 demonstra alguns exemplos de como diferentes tipos de mutações, dentro e fora da região codificadora, podem afetar o mRNA e a proteína.

É importante ter em mente a distinção entre a ocorrência de uma mutação gênica – ou seja, uma alteração na sequência de DNA de um determinado gene – e a detecção de tal evento no nível fenotípico. Muitas mutações de ponto dentro das sequências não codificadoras evocam pouca ou nenhuma alteração fenotípica; essas mutações estão localizadas entre os sítios de ligação ao DNA para proteínas reguladoras. Tais sítios podem ser funcionalmente irrelevantes ou outros sítios dentro do gene podem duplicar a sua função.

16.2 Base molecular das mutações espontâneas

As mutações gênicas podem surgir espontaneamente ou ser induzidas. As **mutações espontâneas** são mutações de ocorrência natural e surgem em todas as células. As **mutações induzidas** surgem por meio da ação de determinados agentes denominados **mutágenos**, que aumentam a taxa na qual as mutações ocorrem. Nesta seção, consideraremos a natureza das mutações espontâneas.

Teste de flutuação de Luria e Delbrück

A origem da alteração hereditária espontânea sempre foi um tópico de interesse

considerável. Entre as primeiras questões indagadas pelos geneticistas estava: "As mutações espontâneas ocorrem em resposta ao agente selecionado ou são variantes presentes a uma baixa frequência na maior parte das populações?" Um sistema experimental ideal para abordar essa questão importante foi a análise de mutações em bactérias que conferem resistência a agentes ambientais específicos normalmente não tolerados por células do tipo selvagem.

Um experimento de Salvador Luria e Max Delbrück, em 1943, foi particularmente influente na moldagem da nossa compreensão sobre a natureza da mutação, não apenas em bactérias, mas nos organismos em geral. Naquela ocasião sabia-se que, se bactérias *E. coli* são espalhadas sobre uma placa de meio nutriente na presença do fago T1, os fagos logo infectam e matam as bactérias. Raramente, porém de modo regular, eram observadas colônias resistentes ao ataque de fagos; essas colônias eram estáveis e, assim, aparentavam ser mutantes genuínos. Entretanto, não se sabia se esses mutantes eram produzidos espontaneamente, porém de modo aleatório no tempo, ou se a presença do fago induzia uma alteração fisiológica que causava resistência.

Luria raciocinou que, se as mutações ocorriam espontaneamente, então poderiam ocorrer em ocasiões diferentes em culturas diferentes. Nesse caso, o número de colônias resistentes por cultura deveria demonstrar alta variação (ou "flutuação", em suas palavras). Posteriormente, ele alegou que a ideia passou por sua mente quando observou os retornos flutuantes obtidos por colegas que apostavam em um caça-níquel do baile de faculdade de um *country club* local; daí, a origem do termo *mutação jackpot*.

Luria e Delbrück projetaram seu **"teste de flutuação"** como se segue. Eles inocularam 20 pequenas culturas, cada uma com algumas células, e as incubaram até que houvesse 10^8 células por mililitro. Ao mesmo tempo, também foi inoculada uma cultura muito maior e ela foi incubada até que houvesse 10^8 células por mililitro. As 20 culturas individuais e 20 amostras de mesmo tamanho da cultura grande foram plaqueadas na presença do fago. As 20 culturas individuais demonstraram alta variação no número de colônias resistentes: 11 placas apresentaram 0 colônia resistente e as remanescentes apresentaram 1, 1, 3, 5, 5, 6, 35, 64 e 107 por placa (Figura 16.5 A). As 20 amostras da cultura grande

demonstraram muito menos variação placa a placa, todas na faixa de 14 a 26. Se o fago estivesse induzindo mutações, não haveria motivo para a flutuação ser mais alta nas culturas individuais, tendo em vista que todas foram expostas ao fago de modo semelhante. A melhor explicação era que a mutação estava ocorrendo aleatoriamente no tempo: as mutações iniciais proporcionaram números mais altos de células resistentes, pois as células mutantes tiveram tempo para produzir muitos descendentes resistentes. As últimas mutações produziram menos células resistentes (Figura 16.5 B). Esse resultado levou ao "paradigma" da mutação; isso é, seja em vírus, bactérias ou eucariotos, podem ocorrer mutações em qualquer célula em qualquer ocasião e a sua ocorrência é aleatória. Por esse e por outro trabalho, Luria e Delbrück receberam o Prêmio Nobel de Fisiologia ou Medicina em 1969. Curiosamente, isso ocorreu após o primeiro aluno de pós-graduação de Luria, James Watson, ter recebido o seu Prêmio Nobel (com Francis Crick em 1964) pela descoberta da estrutura de dupla-hélice do DNA.

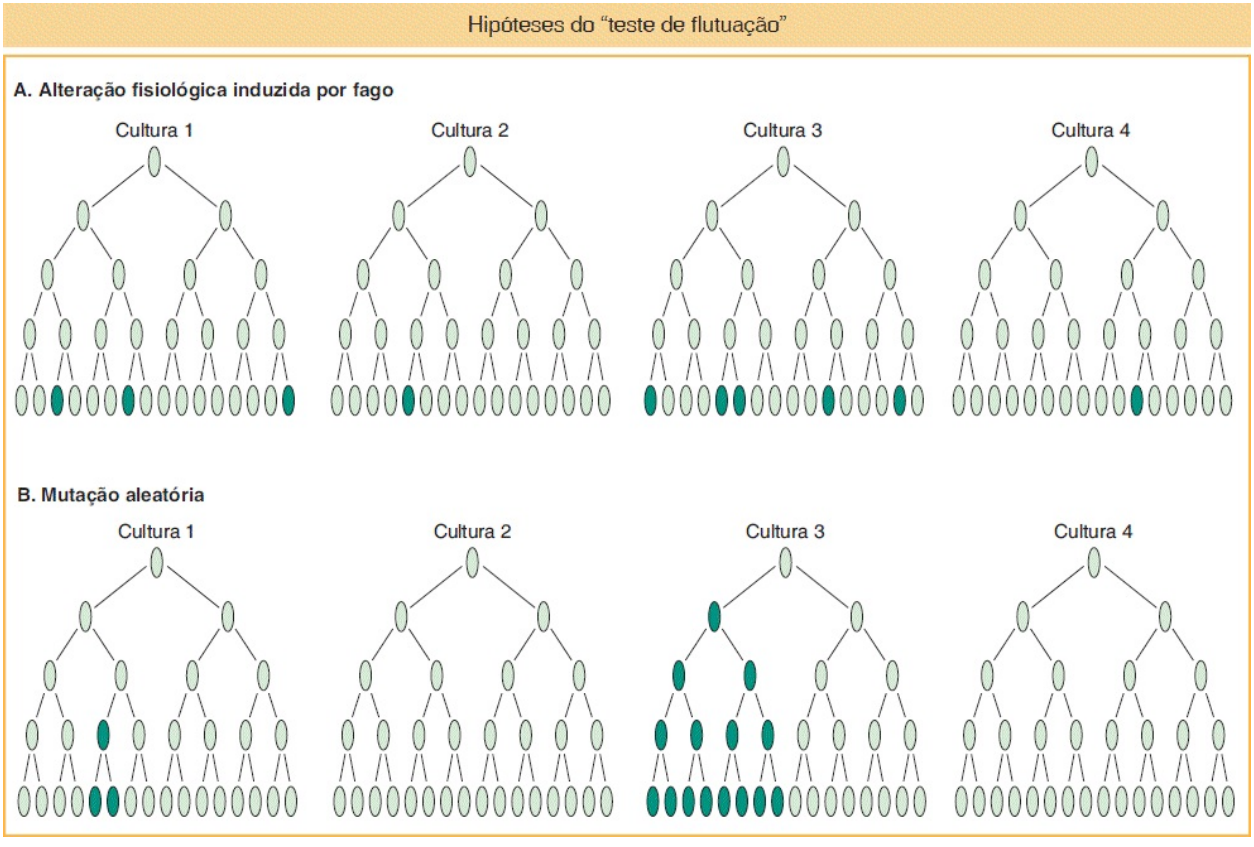


FIGURA 16.5 Estes heredogramas celulares ilustram as expectativas de duas hipóteses contrastantes a

respeito da origem de células resistentes.

Essa análise elegante sugere que as células resistentes são selecionadas pelo agente ambiental (aqui, o fago), em vez de produzidas por ele. A existência de mutantes em uma população pode ser demonstrada diretamente antes da seleção? Essa demonstração foi possibilitada por meio da utilização de uma técnica denominada **plaqueamento em réplica**, desenvolvida em grande parte por Esther Lederberg em 1952. Uma população de bactérias foi plaqueada em meio não seletivo – ou seja, um meio não contendo fagos – e cresceu uma colônia a partir de cada célula. Essa placa foi denominada *placa máster*. Um pedaço de veludo estéril foi pressionado levemente sobre a superfície da placa máster e o veludo coletou células em todos os locais onde havia uma colônia (Figura 16.6). Desse modo, o veludo coletou uma "impressão" das colônias da placa inteira. Em seguida, placas-réplica contendo meio seletivo (ou seja, contendo o fago T1) foram tocadas com o veludo. Ao toque das placas com o veludo, as células contidas nele foram inoculadas nas placas-réplica nas mesmas posições relativas daquelas das colônias na placa máster original. Conforme esperado, foram observadas colônias mutantes resistentes raras nas placas-réplica, mas as múltiplas placas-réplica demonstraram padrões idênticos de colônias resistentes (Figura 16.7). Se as mutações houvessem ocorrido *após* a exposição aos agentes seletivos, os padrões em relação a cada placa teriam sido tão aleatórios quanto as próprias mutações. Os eventos de mutação devem ter ocorrido antes da exposição ao agente seletivo. Novamente, esses resultados confirmam que a mutação está ocorrendo aleatoriamente o tempo todo, em vez de em resposta a um agente seletivo.

> CONCEITO-CHAVE A mutação é um processo aleatório. Qualquer alelo em qualquer célula pode sofrer mutação em qualquer ocasião.

Mecanismos de mutações espontâneas

As mutações espontâneas têm origem a partir de uma diversidade de fontes. Uma

fonte é o processo de replicação do DNA. Embora a replicação do DNA seja um processo notavelmente preciso, são produzidos erros na cópia de milhões e até mesmo bilhões de pares de bases em um genoma. As mutações espontâneas também surgem em parte porque o DNA é uma molécula muito frágil e o próprio ambiente celular pode danificá-lo. Conforme descrito no Capítulo 15, as mutações podem ser causadas até mesmo pela inserção de um elemento de transposição de outro local do genoma. Neste capítulo, enfocamos as mutações que não são causadas por elementos de transposição.

Erros na replicação do DNA. Pode resultar um erro na replicação do DNA quando um par incorreto de nucleotídios (digamos, A-C) é formado na síntese de DNA, levando a uma substituição de bases que pode ser uma transição ou uma transversão. Outros erros podem adicionar ou subtrair pares de bases, de modo que é criada uma mudança de matriz de leitura.

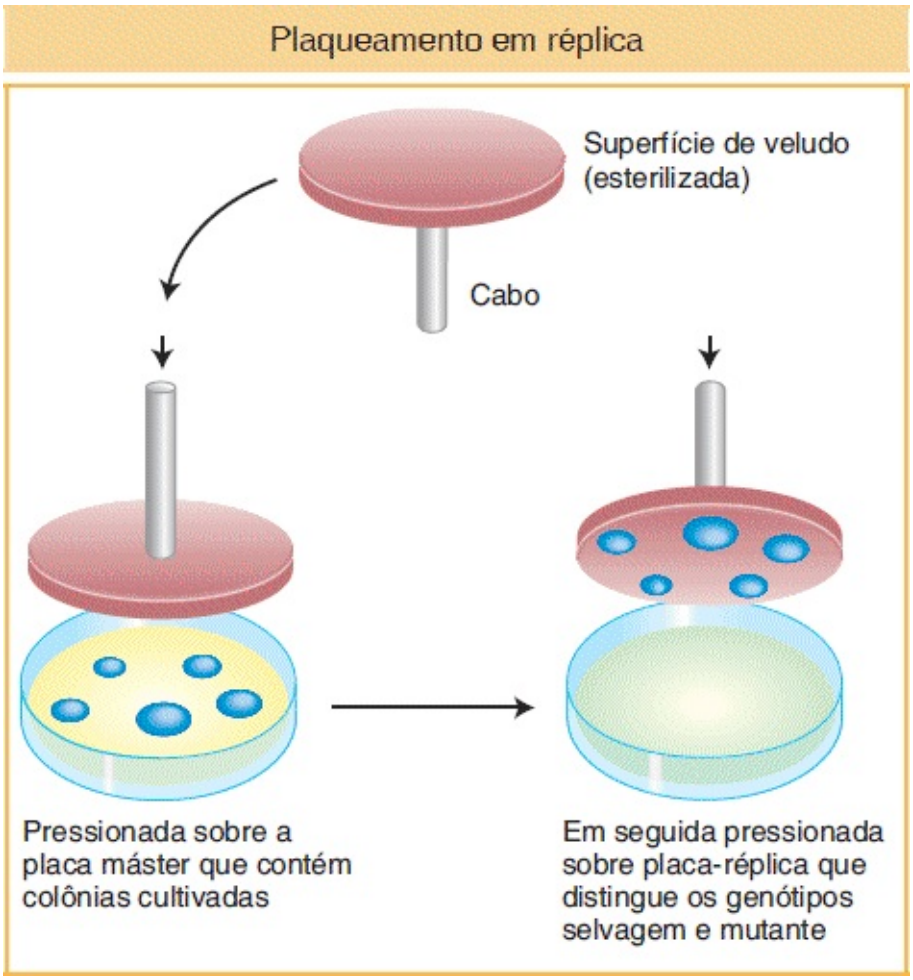


FIGURA 16.6 O plaqueamento em réplica revela colônias mutantes em uma placa máster por meio do seu comportamento em placas-réplica seletivas.

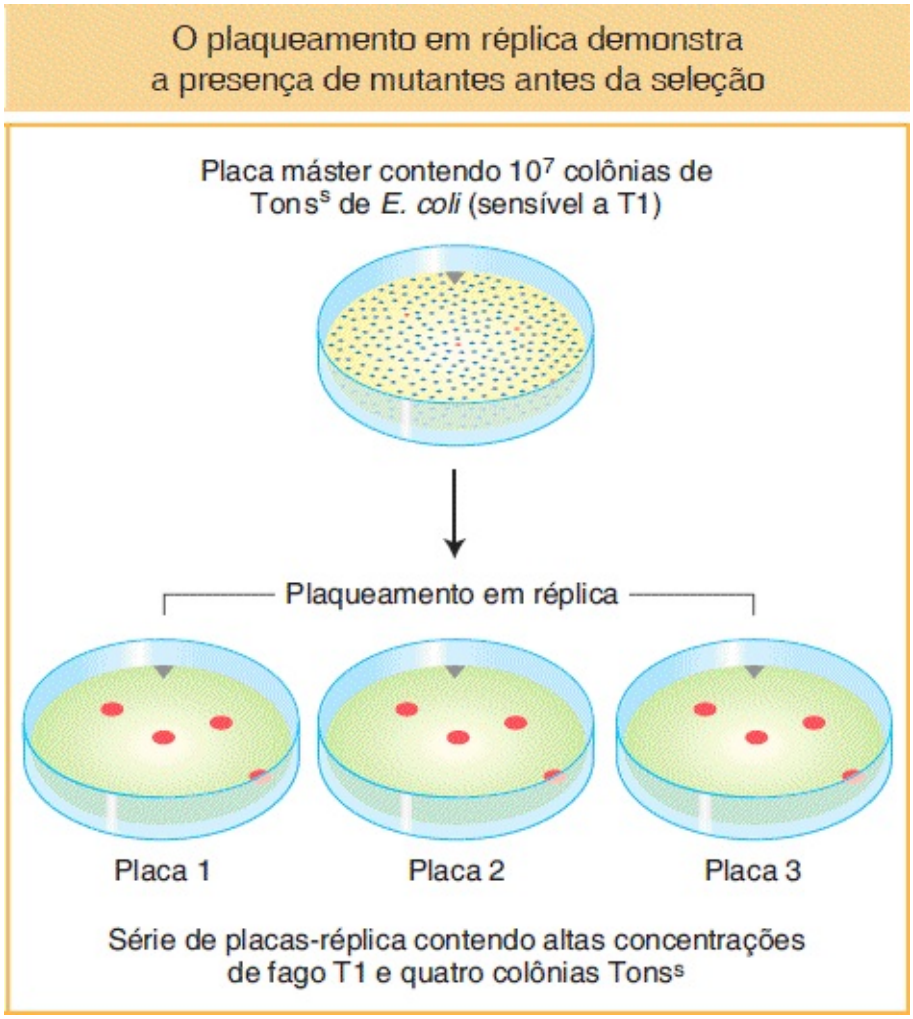


FIGURA 16.7 Os padrões idênticos nas réplicas demonstram que as colônias resistentes são da placa máster.

Transições. Você viu, no Capítulo 7, que cada uma das bases no DNA pode estar presente em uma de diversas formas tautoméricas, que podem parear com a base errada. Malpareamentos também podem resultar quando uma das bases se torna *ionizada*. Esse tipo de pareamento incorreto pode ocorrer mais frequentemente do que os malpareamentos decorrentes de tautomerização. Esses erros com frequência são corrigidos pela função de revisão (edição) da DNA pol III bacteriana (ver Figura 7.18). Se a revisão não ocorrer, todos os malpareamentos descritos até então levam a mutações de transição, nas quais uma purina substitui uma purina ou uma pirimidina substitui uma pirimidina (ver Figura 16.2). Outros sistemas de reparo (descritos posteriormente neste capítulo) corrigem muitas das bases malpareadas que escapam à correção pela função de edição da polimerase.

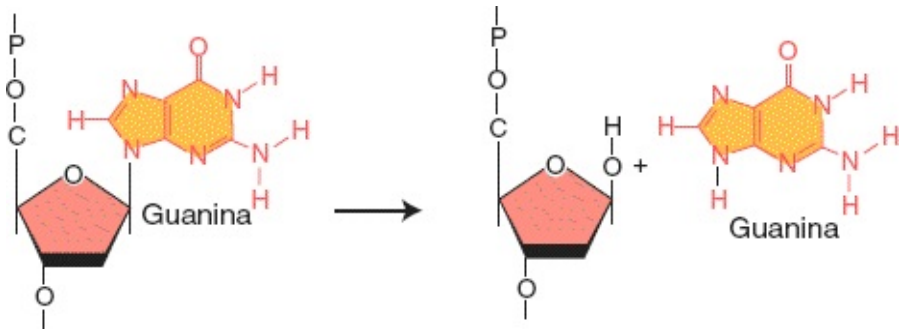
Transversões. Em mutações por transversão, uma pirimidina substitui uma purina ou vice-versa (ver Figura 16.2). A criação de uma transversão por um erro de replicação necessitaria, em algum ponto no curso da replicação, o pareamento errôneo de uma purina com uma purina ou de uma pirimidina com uma pirimidina. Embora as dimensões da dupla-hélice de DNA tornem tais malpareamentos energeticamente desfavoráveis, agora sabemos, a partir de estudos de difração de raios X, que podem ser formados pares G-A, bem como outros pares de purinas com purinas.

Mudanças de matriz de leitura. Os erros de replicação também podem levar a mudanças de matriz de leitura. Relembre, do Capítulo 9, que as referidas mutações resultam em proteínas muito alteradas.

Determinados tipos de erros de replicação podem levar a **mutações indel** – ou seja, inserções ou deleções de um ou mais pares de bases, que produzem mudanças de matriz de leitura quando adicionam ou subtraem uma quantidade de bases não divisível por três (o tamanho de um códon) nas regiões codificadoras de proteína. O modelo predominante (Figura 16.8) propõe que as mutações indel surgem quando alças nas regiões unifilamentares são estabilizadas por "pareamento deslocado" (*slipped*) de sequências repetidas no curso da replicação. Esse mecanismo por vezes é denominado *replicação deslocada*.

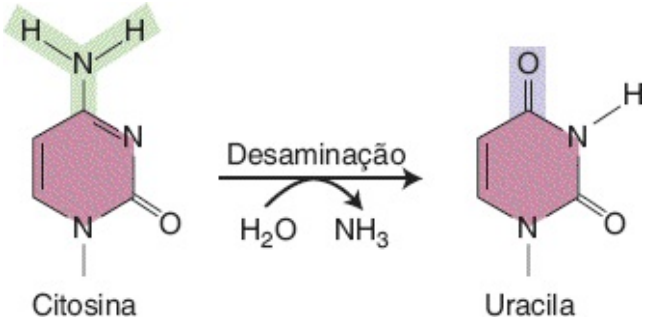
Lesões espontâneas. Além dos erros de replicação, **lesões espontâneas**, danos no DNA de ocorrência natural, podem gerar mutações. Duas das lesões espontâneas mais frequentes resultam da depurinação e desaminação.

A depurinação, a mais comum das duas, é a perda de uma base purina. A depurinação consiste na interrupção da ligação glicosídica entre a base e a desoxirribose e a subsequente perda de uma guanina ou adenina do DNA. O arcabouço do DNA permanece intacto.



Uma célula de mamífero perde de modo espontâneo aproximadamente 10.000 purinas de seu DNA em um período de ciclo celular de 20 horas a 37°C. Se essas lesões persistissem, elas resultariam em um dano genético significativo, tendo em vista que, na replicação, os **sítios apurínicos** resultantes não conseguem especificar uma base complementar à purina original. Entretanto, conforme veremos posteriormente no capítulo, sistemas de reparo eficientes removem os sítios apurínicos. Sob determinadas condições (descritas posteriormente), uma base pode ser inserida através de um sítio apurínico; essa inserção com frequência resultará em uma mutação.

A desaminação da citosina produz uracila.



Os resíduos de uracila não reparados irão parear-se com adenina na replicação, resultando na conversão de um par G · C em um par A · T (uma transição G · C → A · T).

Bases danificadas de modo oxidativo representam um terceiro tipo de lesão espontânea que gera mutações. Espécies ativas de oxigênio, tais como radicais superóxido ($O_2 \cdot^-$), peróxido de hidrogênio (H_2O_2) e radicais hidroxila (· OH), são subprodutos do metabolismo aeróbico normal. Elas podem causar danos

oxidativos ao DNA, bem como nos precursores do DNA (como GTP), resultando em mutação. As mutações por danos oxidativos foram implicadas em uma diversidade de doenças humanas. A Figura 16.9 demonstra dois produtos de dano oxidativo. O produto da 8-oxo-7-hidrodesoxiguanosina (8-oxo dG ou GO) com frequência induz pareamento errôneo com A, resultando em um alto nível de transversões G · T. O produto timidina glicol bloqueia a replicação do DNA, se não reparado.

CONCEITO-CHAVE Mutações espontâneas podem ser geradas por meio de diferentes processos. Erros de replicação e lesões espontâneas geram a maior parte das substituições de bases espontâneas. Os erros de replicação também podem causar deleções que levam a mudanças de matriz de leitura.

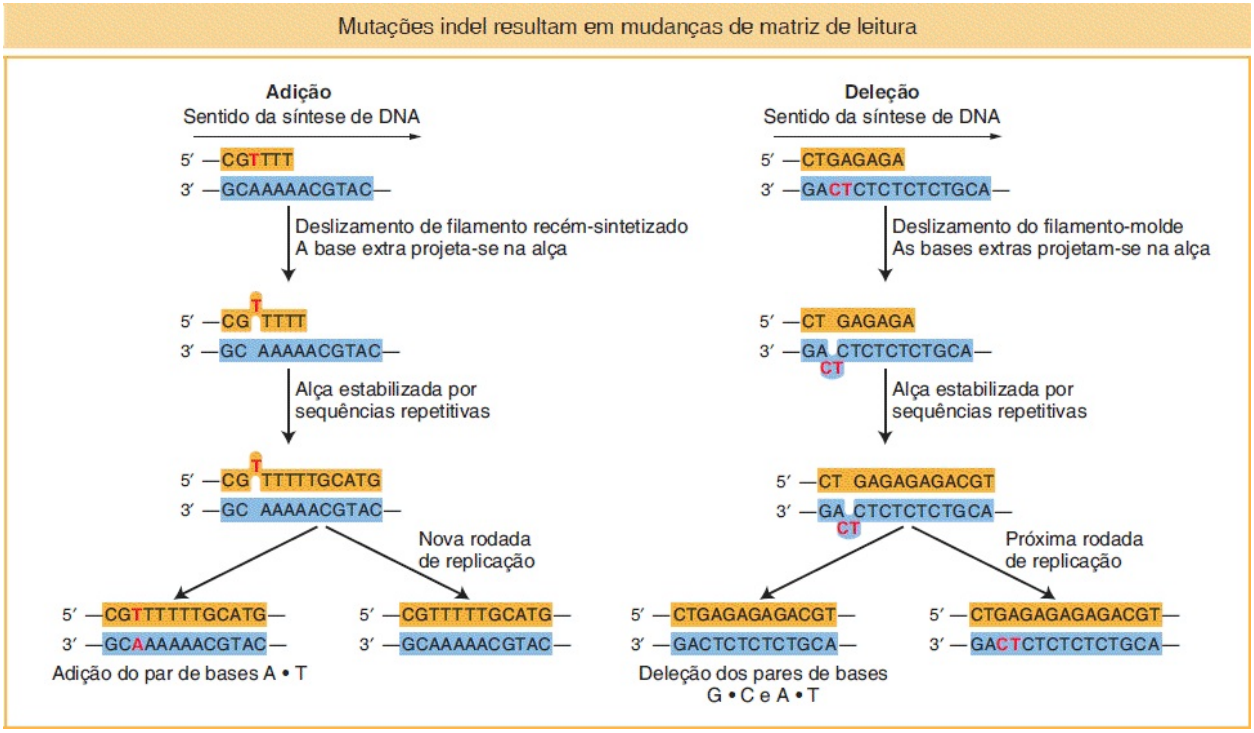


FIGURA 16.8 Inserções e deleções de bases (mutações indel) causam mudanças de matriz de leitura por meio do pareamento errôneo de sequências repetidas no curso da replicação.

Mutações espontâneas em seres humanos | Doenças por repetições de

trinucleotídios

A análise da sequência de DNA revelou as mutações gênicas que contribuem para diversas doenças hereditárias humanas. Muitas são, como esperado, substituições de bases ou tipo indel de um único par de bases. Entretanto, algumas mutações são mais complexas. Uma diversidade desses distúrbios humanos ocorre em virtude da duplicação de sequências curtas repetidas.

Um mecanismo comum responsável por uma diversidade de doenças genéticas é a expansão de uma repetição de três pares de bases. Por esse motivo, elas são denominadas doenças por **repetição de trinucleotídios**. Um exemplo é a doença humana denominada *síndrome do X frágil*. Essa doença é a forma mais comum de comprometimento mental hereditário, que ocorre em aproximadamente 1 de 1.500 homens e 1 de 2.500 mulheres. Ela se manifesta citologicamente por um sítio frágil no cromossomo X, que resulta em quebras *in vitro* (mas isso não leva ao fenótipo da doença). A síndrome do X frágil resulta de mudanças no número de uma repetição de $(CGG)_n$ em uma região do gene *FMR-1* que é transcrita, mas não traduzida (Figura 16.10 A).

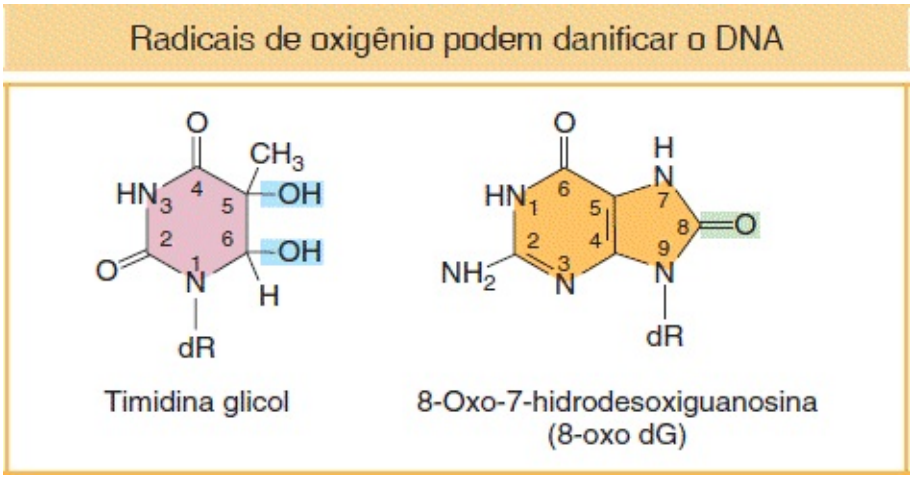


FIGURA 16.9 Produtos formados após o DNA ter sido atacado por radicais de oxigênio. Abreviação: dR = Desoxirribose.

Como o número de repetições se correlaciona ao fenótipo da doença? Os seres humanos normalmente demonstram variação considerável no número de repetições CGG no gene *FMR-1*, que variam de 6 a 54, com o alelo mais

frequente contendo 29 repetições. Por vezes, genitores e avós não afetados dão origem a diversos descendentes com a síndrome do X frágil. Os descendentes com os sintomas da doença apresentam enorme número de repetições, que variam de 200 a 1.300 (Figura 16.10 B). Também se observou que os genitores e os avós não afetados contêm um aumento no número de cópias da repetição, mas que variam de apenas 50 a 200. Por esse motivo, tem sido dito que esses ancestrais carreiam *pré-mutações*. As repetições nesses alelos pré-mutados não são suficientes para causar o fenótipo da doença, mas são muito mais instáveis (ou seja, facilmente expandidas) do que os alelos normais e, assim, levam a uma expansão ainda maior na sua descendência. (Em geral, quanto mais expandida a repetição, maior a instabilidade parece ser.)

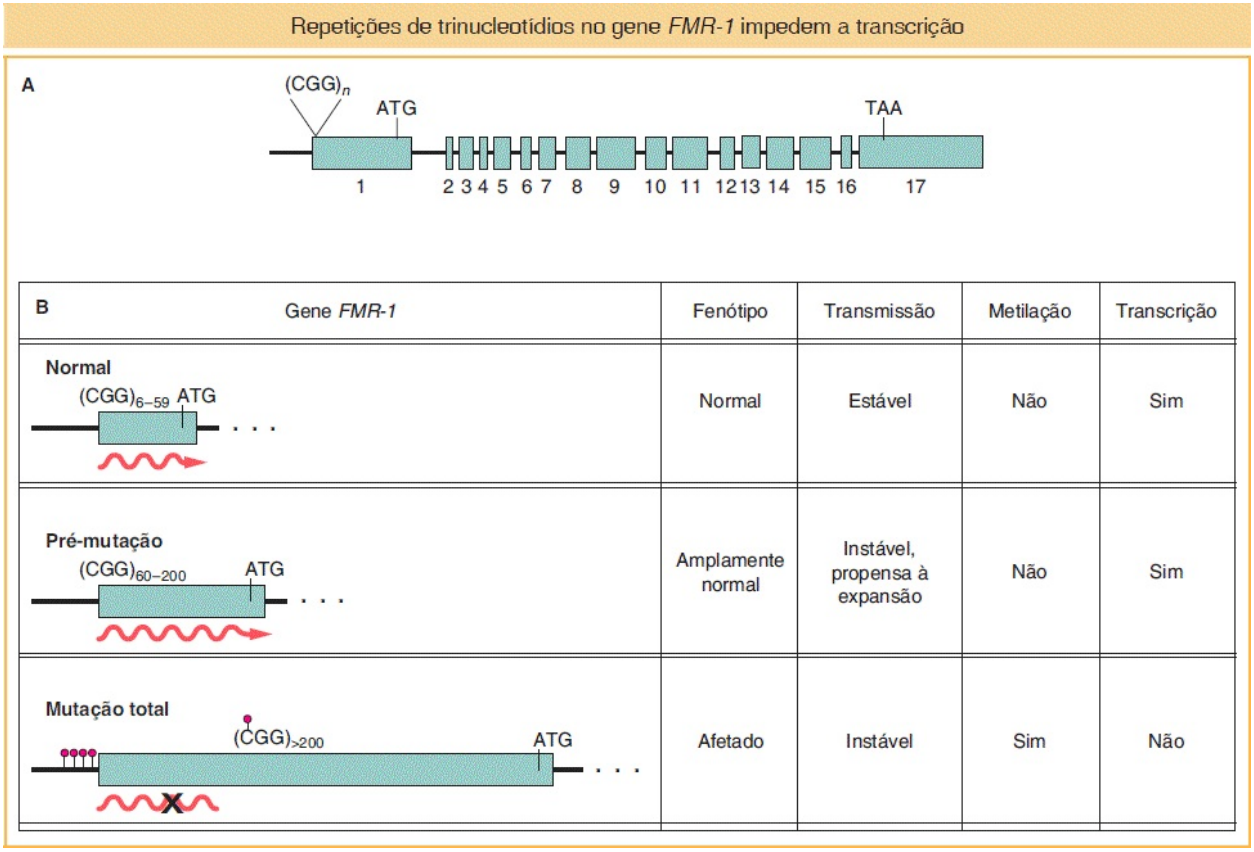


FIGURA 16.10 O gene *FMR-1* na síndrome do X frágil. **A.** Estrutura do éxon e repetição CGG *upstream*. **B.** Transcrição e metilação em alelos normais, pré-mutação e mutação total. Os *círculos vermelhos* representam os grupos metil. (Dados de W. T. O'Donnell e S. T. Warren, *Annu. Rev. Neurosci.* 25, 2002, 315-338, Fig. 1.)

O mecanismo proposto em relação à geração dessas repetições é o deslize na replicação, que ocorre no período da síntese de DNA (Figura 16.11). Entretanto, a frequência extraordinariamente alta de mutação nas repetições de trinucleotídios na síndrome do X frágil sugere que em células humanas, após um limiar de aproximadamente 50 repetições, o maquinário de replicação não consegue replicar fielmente a sequência correta, e ocorrem grandes variações no número de repetições.

Outras doenças, tais como a doença de Huntington (ver Capítulo 2), também foram associadas à expansão das repetições de trinucleotídios em um gene ou em suas regiões reguladoras. Diversos temas gerais se aplicam a essas doenças. Na doença de Huntington, por exemplo, o gene do tipo selvagem *HD* inclui uma sequência repetida, com frequência na região codificadora da proteína, e a mutação está correlacionada com uma expansão considerável dessa região repetida. A gravidade da doença está correlacionada com o número de cópias repetidas.

A doença de Huntington e a doença de Kennedy (também denominada *atrofia muscular bulbar e espinal ligada ao X*) resultam da amplificação de uma repetição de três pares de bases, CAG. Pessoas não afetadas apresentam uma média de 19 a 21 repetições CAG, enquanto os pacientes afetados apresentam uma média de aproximadamente 46. Na doença de Kennedy, que é caracterizada por fraqueza e atrofia muscular progressiva, a expansão da repetição de trinucleotídios está no gene que codifica o receptor de andrógeno.

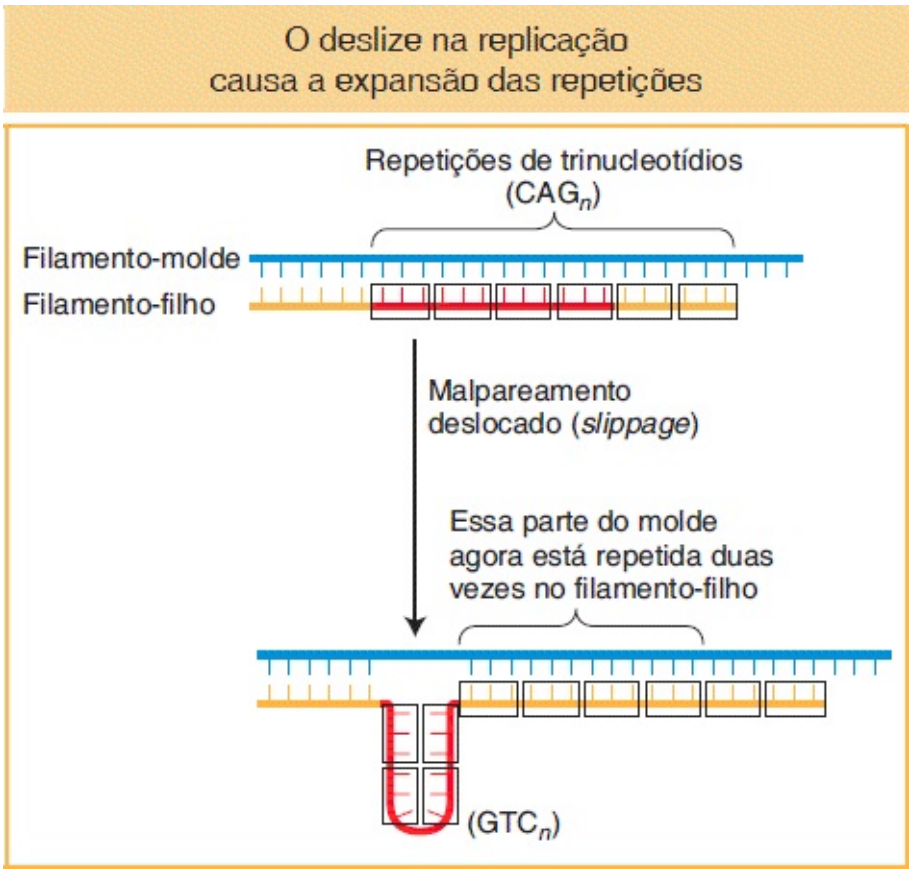


FIGURA 16.11 As regiões de repetições de trinucleotídios são propensas a malpareamento durante a replicação (*alça vermelha*). Consequentemente, a mesma região de repetições de trinucleotídios pode ser duplicada duas vezes no curso da replicação.

As propriedades comuns a algumas doenças por repetição de trinucleotídios sugerem um mecanismo comum por meio do qual os fenótipos anormais são produzidos. Primeiramente, muitas dessas doenças aparentam incluir neurodegeneração – ou seja, morte celular dentro do sistema nervoso central. Em segundo lugar, nas referidas doenças, as repetições de trinucleotídios ocorrem nas matrizes de leitura abertas dos transcritos do gene mutado, levando a expansões ou contrações do número de repetições de um único aminoácido no polipeptídio (p. ex., repetições de CAG codificam uma repetição de poliglutamina). Desse modo, é fácil compreender o motivo de essas doenças resultarem de expansões de unidades do tamanho de códons com o comprimento de três pares de bases.

Entretanto, essa explicação não é válida para todas as doenças por repetição de trinucleotídios. Afinal, na síndrome do X frágil, a expansão de trinucleotídios está

próxima da extremidade de 5′ do mRNA do *FMR-1*, *antes* do sítio de início da tradução. Portanto, não podemos atribuir as anormalidades fenotípicas das mutações no *FMR-1* a um efeito sobre a estrutura proteica. Uma pista para o problema com os genes *FMR-1* mutantes é que, contrariamente ao gene normal, eles são hipermetilados, uma característica associada aos genes transcricionalmente silenciados (ver Figura 16.10 B). Com base nesses achados, supõe-se que a repetição expandida leva a alterações na estrutura da cromatina que silenciam a transcrição gênica mutante (ver Capítulo 12). Em amparo a esse modelo está o achado de que o gene *FMR-1* está deletado em alguns pacientes com síndrome do X frágil. Essas observações amparam uma mutação de perda de função.

> CONCEITO-CHAVE Doenças por repetição de trinucleotídios surgem por meio da expansão do número de cópias de uma sequência de três pares de bases normalmente presente em diversas cópias, com frequência na região codificadora de um gene.

16.3 Base molecular das mutações induzidas

Enquanto algumas mutações são produzidas espontaneamente dentro da célula, outras fontes de mutação estão presentes no ambiente, aplicadas intencionalmente em laboratório ou encontradas naturalmente na vida diária. A produção de mutações em laboratório por meio da exposição a mutágenos é denominada **mutagênese** e diz-se que o organismo é *mutagenizado*.

Mecanismos de mutagênese

Os mutágenos induzem mutações por meio de no mínimo três mecanismos diferentes. Eles podem *substituir* uma base no DNA, *alterar* uma base de modo que ela realize especificamente o pareamento errôneo com outra base ou *danificar* uma base de modo que ela não consiga mais parear com qualquer base sob condições normais. Induzir mutações em genes e observar as consequências

fenotípicas é uma das estratégias experimentais primárias utilizadas pelos geneticistas.

Incorporação de análogos de bases. Alguns compostos químicos são suficientemente semelhantes às bases nitrogenadas normais do DNA, de modo que ocasionalmente eles são incorporados ao DNA em lugar das bases normais; os referidos compostos são denominados **análogos de bases**. Após o seu posicionamento, esses análogos apresentam propriedades de pareamento diferentes àquelas das bases normais; portanto, eles podem produzir mutações ao causar a inserção de nucleotídios incorretos em oposição a eles na replicação. O análogo de base original existe apenas em um único filamento, mas pode causar uma substituição de par de nucleotídios que é replicada em todas as cópias do DNA descendentes do filamento original.

Um análogo de base amplamente utilizado em pesquisas é a 2-aminopurina (2-AP). Esse análogo da adenina consegue se parear com a timina, mas também consegue realizar o pareamento errôneo com a citosina quando protonado, conforme demonstrado na Figura 16.12. Portanto, quando a 2-AP é incorporada no DNA por meio do pareamento com timina, ela consegue gerar transições A · T → G · C por meio do pareamento errôneo com citosina nas replicações subsequentes. Ou, se 2-AP for incorporada por meio do pareamento errôneo com citosina, então resultarão transições G · C → A · T quando ela parear com timina. Estudos genéticos demonstraram que 2-AP causa quase exclusivamente transições.

Malpareamento específico. Alguns mutágenos não são incorporados ao DNA, mas, em vez disso, alteram uma base, de modo que ela formará um malpareamento específico. Determinados agentes alquilantes, tais como o etilmetanossulfonato (EMS) e a amplamente utilizada nitrosoguanidina (NG), operam por meio dessa via.

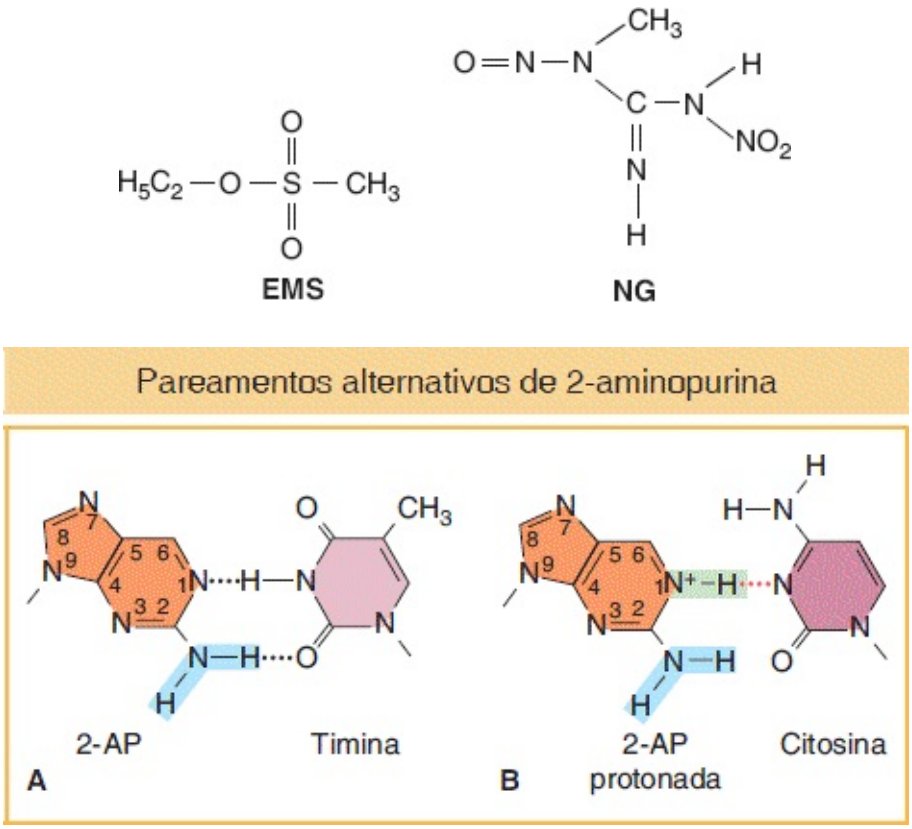


FIGURA 16.12 A. Um análogo da adenina, a 2-aminopurina (2-AP), consegue parear com timina. **B.** Em seu estado protonado, 2-AP consegue parear com citosina.

Tais agentes adicionam grupos alquila (um grupo etil em EMS e um grupo metil em NG) a muitas posições em todas as quatro bases. Entretanto, a formação de uma mutação está mais bem-correlacionada com uma adição ao oxigênio na posição 6 da guanina para criar uma *O*-6-alquilguanina. Essa adição leva ao malpareamento direto com timina, conforme demonstrado na Figura 16.13, que resultará em transições G · C → A · T na próxima rodada de replicação.

Os agentes alquilantes também podem modificar as bases em dNTPs (em que N é qualquer base), os quais são precursores na síntese de DNA.

Agentes intercalares. Os **agentes intercalares** formam outra classe importante de modificadores de DNA. Esse grupo de compostos inclui a proflavina, a acridina laranja e uma classe de substâncias químicas denominadas compostos ICR (Figura 16.14 A). Esses agentes são moléculas planares que mimetizam pares de bases e que conseguem inserir-se (intercalar-se) entre as bases nitrogenadas

empilhadas no cerne da dupla-hélice de DNA (Figura 16.14 B). Nessa posição intercalada, tal agente pode causar uma inserção ou uma deleção de um único par de nucleotídios.

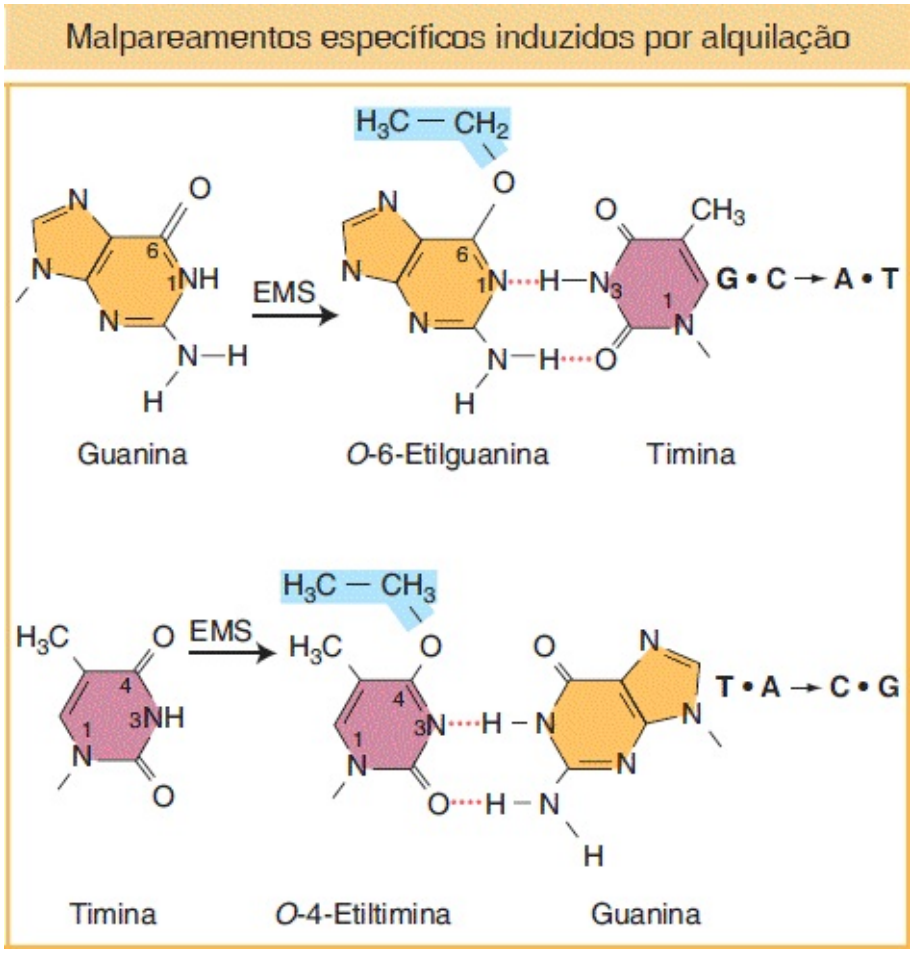


FIGURA 16.13 O tratamento com EMS altera a estrutura da guanina e da timina e leva a malpareamentos.

Danos a bases. Um grande número de mutágenos danifica uma ou mais bases e, assim, nenhum pareamento de bases específico é possível. O resultado é um bloqueio de replicação, tendo em vista que a síntese de DNA não prosseguirá além de uma base que não consegue especificar seu parceiro complementar por meio de pontes de hidrogênio. Os bloqueios de replicação podem causar mutações adicionais – conforme será explicado posteriormente no capítulo (ver seção sobre o reparo por excisão de bases).

A luz ultravioleta normalmente causa danos às bases nucleotídicas na maior

parte dos organismos, pois gera uma diversidade de tipos distintos de alterações no DNA, denominados *fotoprodutos* (da palavra *foto*, em referência à "luz"). Desses produtos, os que mais provavelmente levam a mutações são duas lesões diferentes que unem resíduos de pirimidina adjacentes no mesmo filamento. Esses fotoprodutos são o fotodímero pirimidina ciclobutano e o fotoproduto 6-4 (Figura 16.15).

A radiação ionizante resulta na formação de moléculas ionizadas e excitadas, que podem danificar o DNA. Em virtude da natureza aquosa dos sistemas biológicos, as moléculas geradas pelos efeitos da radiação ionizante sobre a água produzem a maior parte dos danos. São produzidos muitos tipos diferentes de espécies reativas de oxigênio, mas as que causam mais danos às bases do DNA são · OH, O_2^- e H_2O_2. Essas espécies levam à formação de diferentes adutos e produtos de degradação. Entre os mais prevalentes, ilustrados na Figura 16.9, estão a timidina glicol e 8-oxo dG, ambas podendo resultar em mutações.

A radiação ionizante também pode danificar o DNA diretamente, em vez de por meio de espécies reativas de oxigênio. Tal radiação pode causar a quebra da ligação *N*-glicosídica, levando à formação de sítios apurínicos ou apirimidínicos, e pode causar quebras de filamentos. De fato, as quebras de filamentos são responsáveis pela maior parte dos efeitos letais da radiação ionizante.

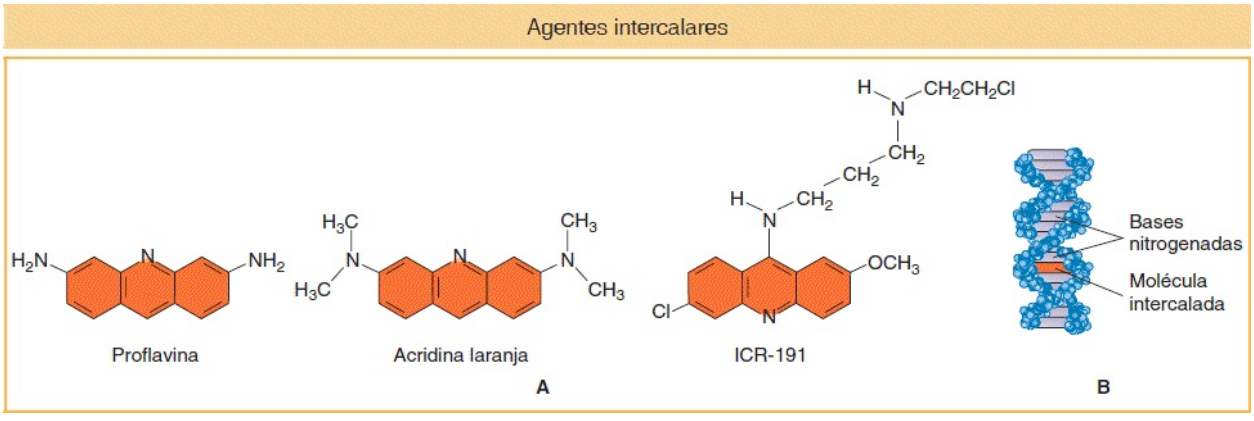


FIGURA 16.14 Estruturas de agentes intercalares comuns (**A**) e sua interação com o DNA (**B**). (Dados de L. S. Lerman, "The Structure of the DNA-Acridine Complex", *Proc. Natl. Acad. Sci. USA* 49, 1963, 94.)

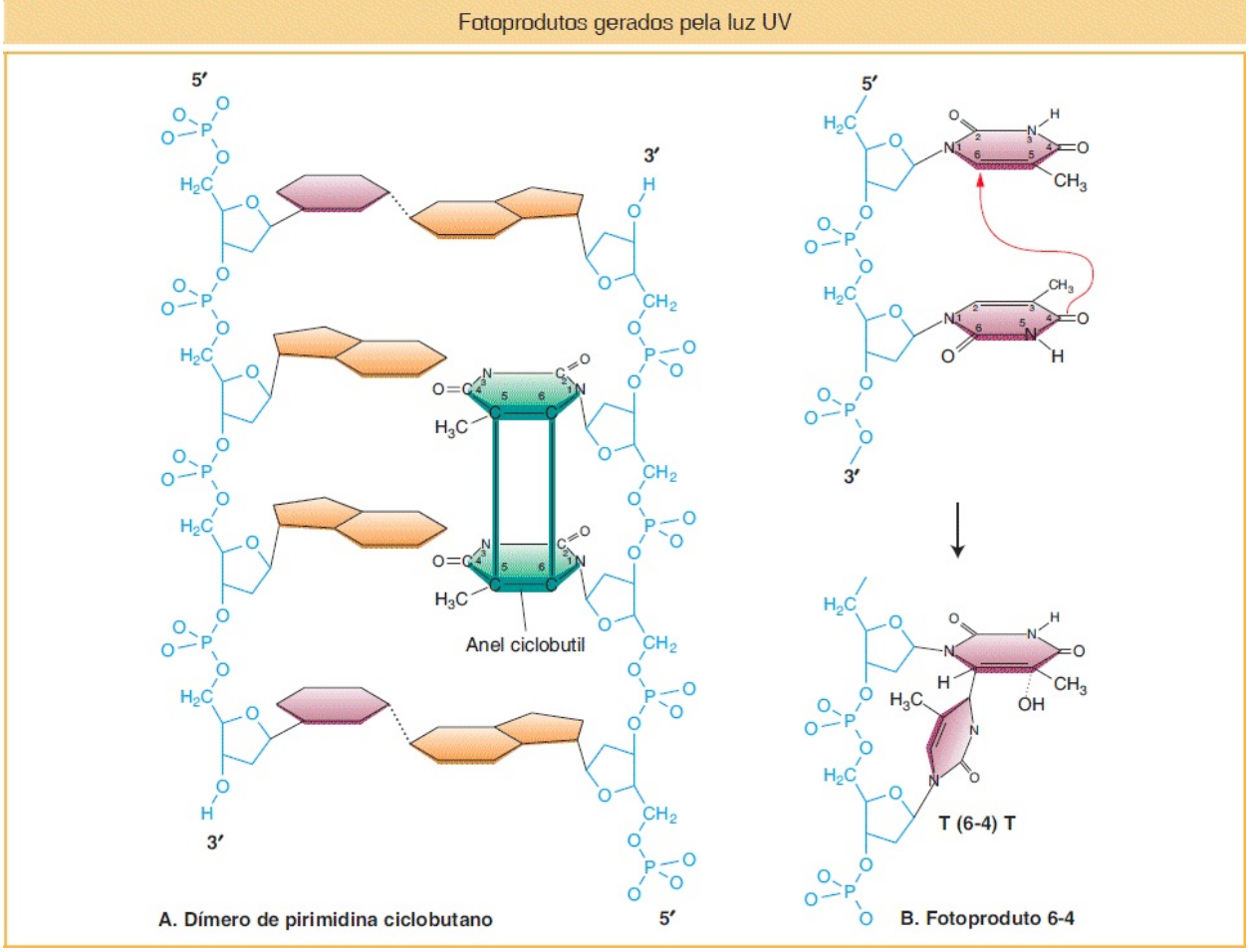


FIGURA 16.15 Os fotoprodutos que unem pirimidinas adjacentes no DNA estão fortemente correlacionados com mutagênese.

A aflatoxina B_1 é um poderoso carcinógeno que se une à guanina na posição N-7 (Figura 16.16). A formação desse produto de adição leva à quebra da ligação entre a base e o açúcar, liberando, assim, a base e gerando um sítio apurínico. A aflatoxina B_1 é um membro de uma classe de carcinógenos químicos conhecidos como produtos de adição volumosos quando eles se ligam de modo covalente ao DNA. Outros exemplos incluem os epóxidos diol de benzo(a)pireno, um composto produzido por mecanismos de combustão interna. Todos os compostos dessa classe induzem mutações, embora por mecanismos nem sempre claros.

> CONCEITO-CHAVE Os mutágenos induzem mutações por meio de uma

diversidade de mecanismos. Alguns mutágenos mimetizam bases normais e são incorporados ao DNA, onde podem parear de modo incorreto. Outros danificam as bases e causam malpareamento específico ou destroem o pareamento, causando o não reconhecimento de bases.

Teste de Ames | Avaliação de mutágenos em nosso ambiente

Foi sintetizado um número enorme de compostos químicos e muitos apresentam possíveis aplicações comerciais. Aprendemos pelo modo mais difícil que os possíveis benefícios dessas aplicações devem ser ponderados com relação aos riscos para a saúde e ambientais. Portanto, é essencial a apresentação de técnicas de triagem eficientes para avaliar alguns dos riscos de um grande número de compostos.

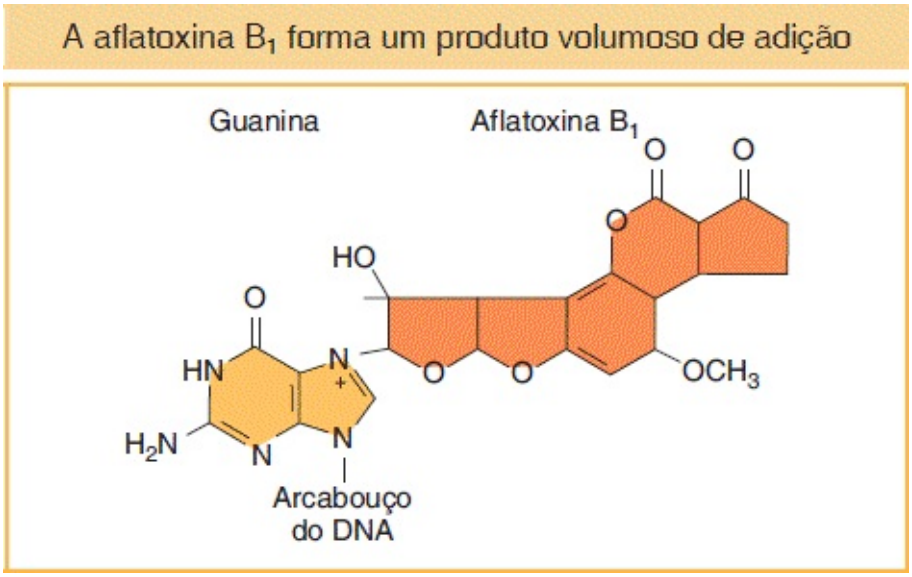


FIGURA 16.16 A aflatoxina B_1 metabolicamente ativada se liga ao DNA.

Muitos compostos são possíveis agentes causadores de câncer (carcinógenos) e, assim, a apresentação de sistemas-modelo válidos, nos quais a carcinogênese dos compostos possa ser avaliada de modo eficiente e efetivo, é muito importante. Entretanto, a utilização de um sistema-modelo de mamífero, tal como o camundongo, é muito lenta, consome muito tempo e é dispendiosa.

Na década de 1970, Bruce Ames reconheceu que existe uma forte correlação entre a capacidade dos compostos de causarem câncer e mutações. Ele conjeturou que a medição das taxas de mutação em sistemas bacterianos seria um modelo eficaz para a avaliação da mutagenicidade dos compostos como um primeiro nível de detecção de possíveis carcinógenos. Entretanto, tornou-se claro que nem todos os carcinógenos eram por si próprios mutagênicos; em vez disso, alguns metabólitos de carcinógenos produzidos no corpo são, de fato, os agentes mutagênicos. Tipicamente, esses metabólitos são produzidos no fígado, e as reações enzimáticas que convertem os carcinógenos em metabólitos bioativos não ocorrem em bactérias.

Ames percebeu que ele poderia superar esse problema por meio do tratamento de linhagens especiais da bactéria *Salmonella typhimurium* com extratos de fígado de rato contendo enzimas metabólicas (Figura 16.17). As linhagens especiais de *S. typhimurium* apresentavam um dos vários alelos mutantes de um gene responsável pela síntese de histidina que sabidamente são conhecidos por "reverter" (ou seja, retornar ao fenótipo do tipo selvagem) apenas por meio de determinados tipos de eventos mutacionais adicionais. Por exemplo, um alelo denominado TA100 poderia ser revertido para o tipo selvagem apenas por mutação de substituição de base, enquanto TA1538 e 1535 poderiam ser revertidos apenas por mutações indel que resultam em uma mudança de matriz de leitura da proteína (Figura 16.18).

As bactérias tratadas de cada uma dessas linhagens foram expostas ao composto do teste e em seguida foram cultivadas em placas de Petri contendo meio sem histidina. A ausência desse nutriente assegurou que apenas os indivíduos revertentes contendo a substituição de base apropriada ou mudança de matriz de leitura cresceriam. O número de colônias em cada placa e o número total de bactérias testadas foram determinados, possibilitando que Ames medisse a frequência da reversão. Os compostos que produziam metabólitos que induziam elevados níveis de reversão em relação aos extratos de fígado de controle não tratados seriam então claramente mutagênicos e seriam possíveis carcinógenos. Portanto, o **teste de Ames** proporcionou um modo importante para triar milhares de compostos e avaliar um aspecto de seu risco para a saúde e o ambiente. Ele

ainda está em uso atualmente como uma importante ferramenta para a avaliação da segurança de compostos químicos.

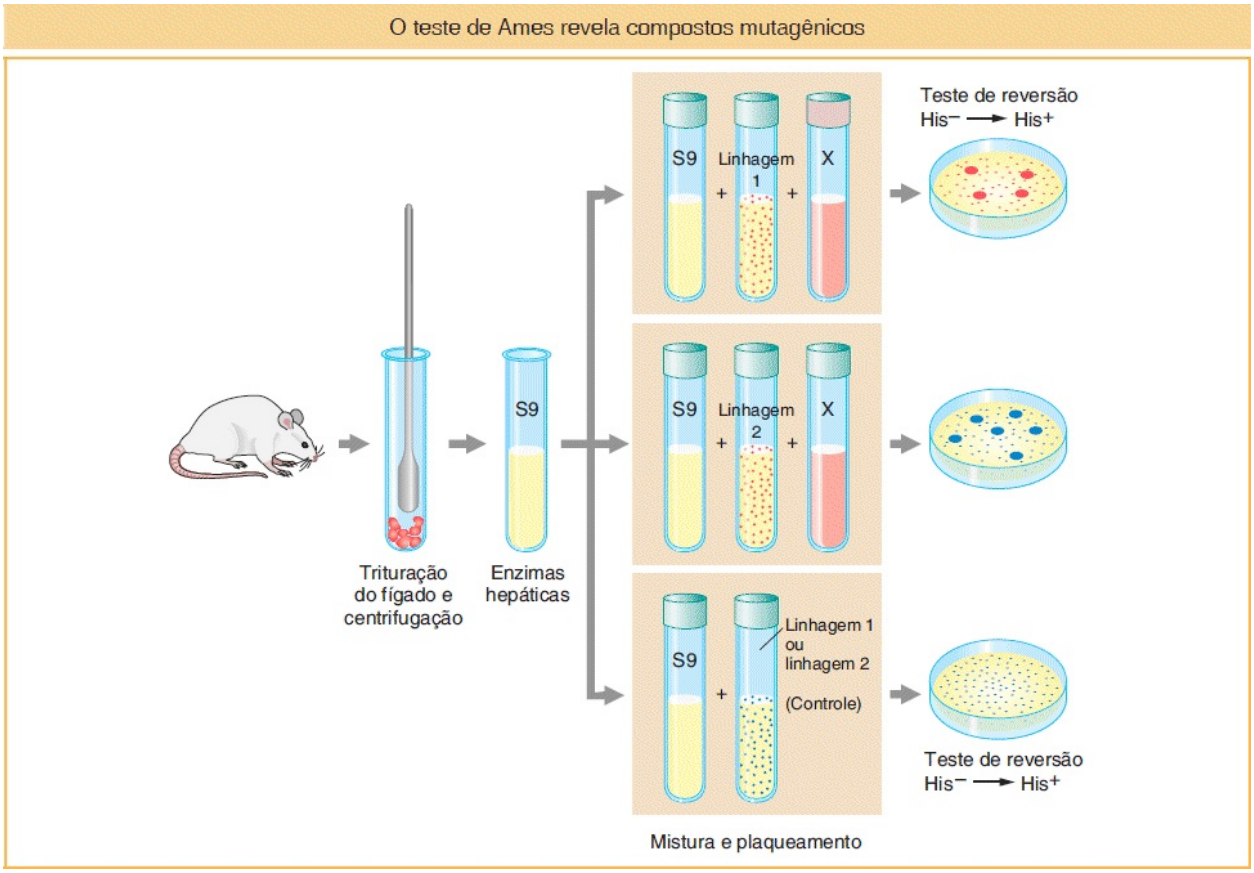


FIGURA 16.17 Resumo do procedimento utilizado no teste de Ames. Enzimas hepáticas solubilizadas (S9) são adicionadas a uma suspensão de bactérias auxotróficas em uma solução do possível carcinógeno (X). A mistura é plaqueada em um meio que não contém histidina. A presença de revertentes indica que a substância química é um mutágeno e possivelmente também um carcinógeno.

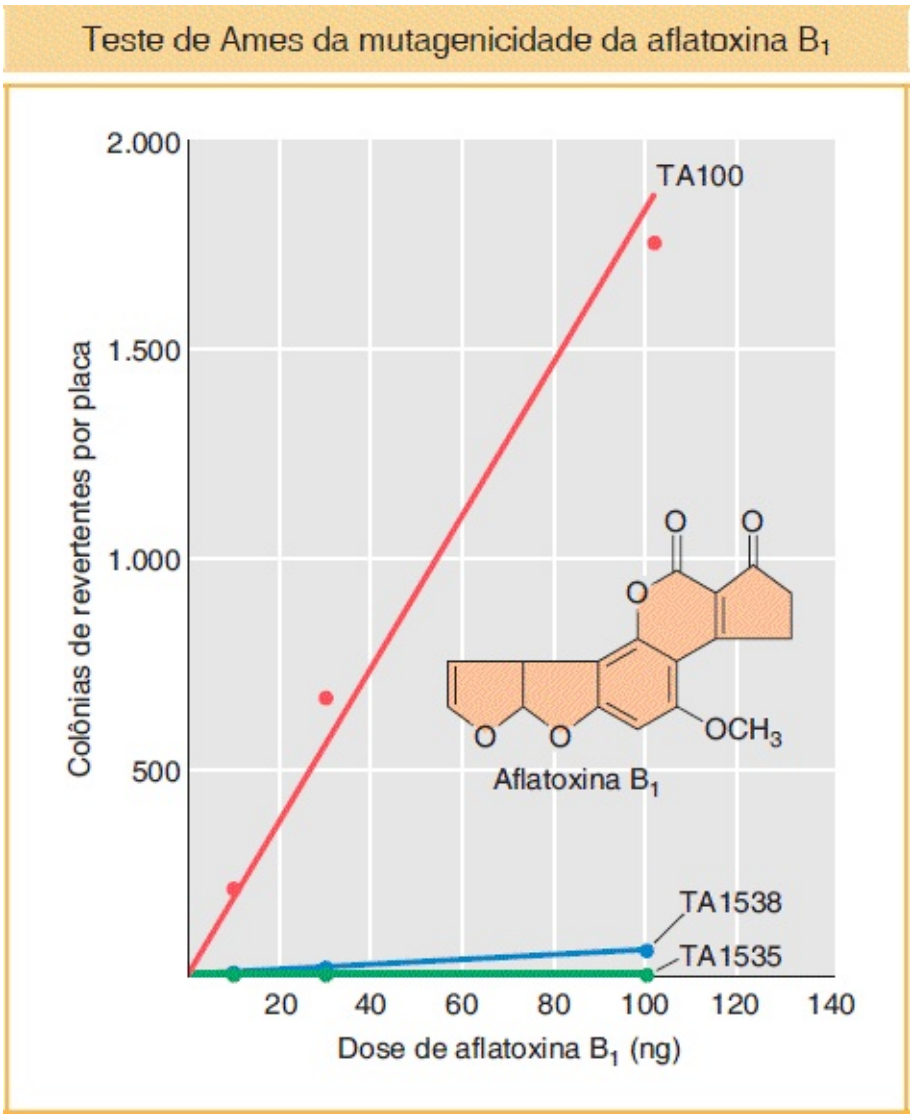


FIGURA 16.18 TA100, TA1538 e TA1535 são linhagens de *Salmonella* que contêm diferentes mutações auxotróficas de histidina. A linhagem TA100 é altamente sensível à reversão por meio da substituição de par de bases. As linhagens TA1535 e TA1538 são sensíveis à reversão por meio da mudança de matriz de leitura. Os resultados do teste demonstram que a aflatoxina B_1 é um mutágeno potente que causa substituições de pares de bases, mas não mudanças de matriz de leitura. (Dados de J. McCann e B. N. Ames, in W. G. Flamm e M. A. Mehlman, eds. *Advances in Modern Technology*, vol. 5. Direitos autorais por Hemisphere Publishing Corporation, Washington, D.C.)

16.4 Mecanismos biológicos de reparo

Após pesquisar os diversos modos por meio dos quais o DNA pode ser danificado – sejam fontes intracelulares (replicação, oxigênio reativo e assim por

diante) ou extracelulares (ambientais: luz UV, radiação ionizante, mutágenos) –, você pode estar ponderando como a vida conseguiu superar esses obstáculos e evoluir durante bilhões de anos. O fato é que os organismos, que variam desde bactérias até seres humanos e plantas, conseguem reparar de modo eficiente seus DNA. Esses organismos utilizam uma diversidade de mecanismos de reparo que, em conjunto, empregam cerca de 100 proteínas conhecidas. De fato, nossa atual compreensão de é que o DNA é a única molécula que os organismos reparam, em vez de substituir. Conforme você verá, a falha desses sistemas de reparo é uma causa significativa das muitas doenças humanas hereditárias.

O mais importante mecanismo de reparo foi mencionado brevemente no Capítulo 7 – a função de revisão das DNA polimerases que replicam o DNA como parte do replissomo. Conforme lá observado, tanto a DNA polimerase I quanto a DNA polimerase III são capazes de realizar a excisão das bases malpareadas inseridas erroneamente. Agora examinaremos algumas das outras vias de reparo, iniciando com o reparo sem erros.

Reversão direta de DNA danificado

O modo mais direto de reparar uma lesão é revertê-la diretamente, regenerando, assim, a base normal (Figura 16.19). Embora a maior parte dos tipos de danos seja essencialmente irreversível, as lesões podem ser reparadas por meio de reversão direta em alguns casos. Um caso é um fotodímero mutagênico causado pela luz UV. O dímero de pirimidina ciclobutano (CPD) pode ser reparado por uma enzima denominada *CPD fotoliase*. A enzima se liga ao fotodímero e o divide para regenerar as bases originais. Esse mecanismo de reparo é denominado *fotorreativação*, tendo em vista que a enzima necessita de luz para atuar. São necessárias outras vias de reparo para remover danos de UV na ausência de luz do comprimento de onda apropriado (> 300 nm).

Alquiltransferases são enzimas que também revertem as lesões diretamente. Elas removem determinados grupos alquila que foram adicionados à posição *O*-6 da guanina (ver Figura 16.13) por mutágenos tais como nitrosoguanidina e etilmetanossulfonato. A metiltransferase de *E. coli* tem sido bem estudada. Essa enzima transfere o grupo metil da *O*-6-metilguanina para um resíduo de cisteína

no sítio ativo da enzima. Entretanto, a transferência inativa a enzima e, assim, esse sistema de reparo pode ser saturado se o nível de alquilação for suficientemente alto.

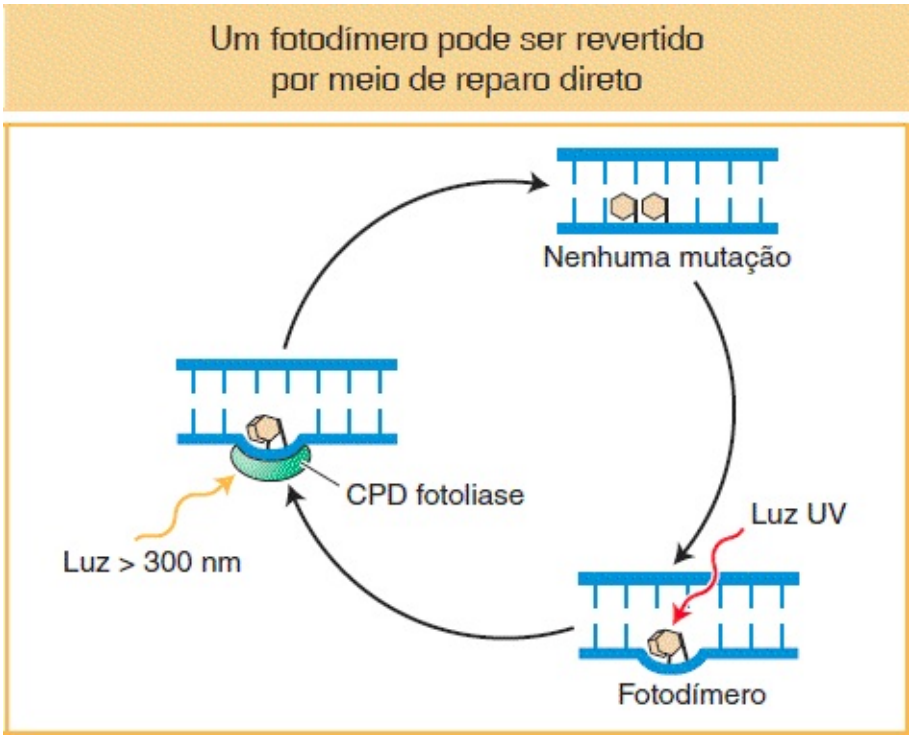


FIGURA 16.19 A enzima CPD fotoliase divide um fotodímero de pirimidina ciclobutano para reparar esta mutação.

Reparo por excisão de base

Um princípio abrangente que orienta os sistemas genéticos celulares é o poder da complementaridade da sequência de nucleotídios. (Lembre que a análise genética também depende fortemente desse princípio.) Sistemas de reparo importantes exploram as propriedades da complementaridade antiparalela para restaurar os segmentos de DNA danificados até o seu estado não danificado inicial. Nesses sistemas, uma base ou um segmento mais longo de uma cadeia de DNA é removido e substituído por um segmento de nucleotídios recém-sintetizado complementar ao filamento-molde oposto. Contrariamente aos exemplos de reversão de danos descritos na seção precedente, essas vias incluem a remoção e a substituição de uma ou mais bases.

O primeiro sistema de reparo desse tipo que examinaremos é o **reparo por excisão de base**. Após a revisão do DNA pela DNA polimerase, o reparo por excisão de base é o mecanismo mais importante utilizado para remover bases incorretas ou danificadas. O principal alvo do reparo por excisão de base é o dano não volumoso a bases. Esse tipo de dano pode resultar da diversidade de causas mencionadas nas seções precedentes, incluindo metilação, desaminação, oxidação ou perda espontânea de uma base do DNA. O reparo por excisão de base (Figura 16.20) é realizado por DNA glicosilases que clivam as ligações base-açúcar, liberando, assim, as bases alteradas e gerando sítios apurínicos ou apirimidínicos (AP). Uma enzima denominada endonuclease AP em seguida corta o filamento danificado *upstream* do sítio AP. Uma terceira enzima, desoxirribofosfodiesterase, limpa o arcabouço ao remover um trecho de resíduos açúcar-fosfato vizinhos, de modo que uma DNA polimerase possa preencher o intervalo com nucleotídios complementares ao outro filamento. A DNA ligase em seguida sela o novo nucleotídio no arcabouço (ver Figura 16.20).

Existem diversas DNA glicosilases. Uma, a uracila-DNA glicosilase, remove a uracila do DNA. As unidades de uracila, que resultam da desaminação espontânea da citosina (ver anteriormente), podem levar a uma transição de C para T, se não forem reparadas. Uma vantagem de ter timina (5-metiluracila) em vez da uracila como parceira natural de pareamento da adenina no DNA é que os eventos de desaminação espontânea da citosina podem ser reconhecidos como anormais e em seguida podem ser excisados e reparados. Se a uracila fosse um constituinte normal do DNA, tal reparo não seria possível.

Entretanto, a desaminação impõe outros problemas para as bactérias e os eucariotos. Ao analisar um grande número de mutações no gene *lacI*, Jeffrey Miller identificou locais no gene nos quais uma ou mais bases estavam propensas a mutações frequentes. Miller observou que esses assim denominados pontos quentes (*hotspots*) mutacionais correspondiam a desaminações em determinados resíduos de citosina. A análise da sequência de DNA de *hotspots* de transição G · C → T · A no gene *lacI* demonstrou que resíduos de 5-metilcitosina estão presentes em cada *hotspot*. Relembre do Capítulo 12 que o DNA eucariótico pode ser metilado para inativar genes. De modo semelhante, *E. coli* e outras

bactérias também metilam o seu DNA, embora para diferentes finalidades. Alguns dos dados desse estudo de *lacI* estão demonstrados na Figura 16.21. A altura de cada barra no gráfico representa a frequência de mutações em cada sítio. Pode-se observar que as posições dos resíduos de 5-metilcitosina estão bem-correlacionadas com os sítios mais mutáveis.

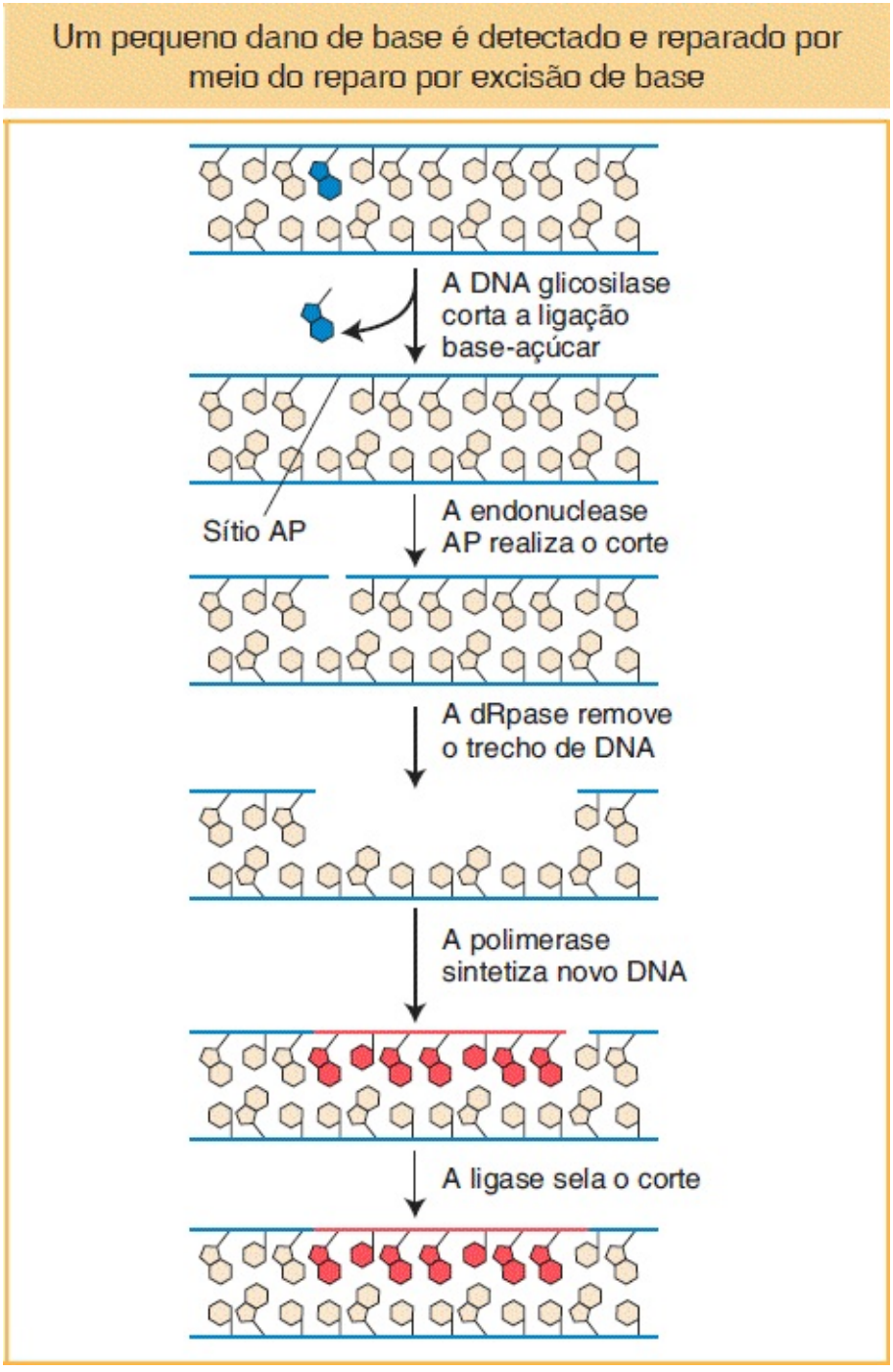
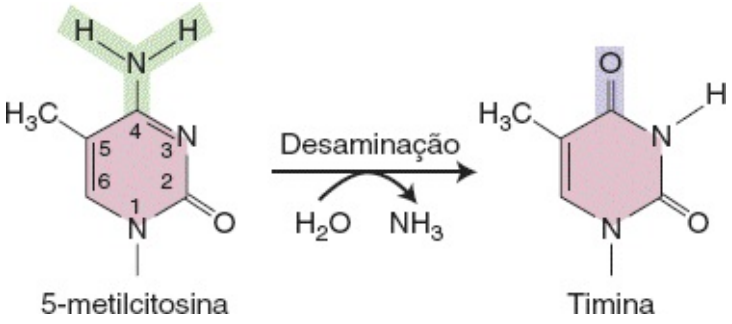


FIGURA 16.20 No reparo por excisão de base, as bases danificadas são removidas e reparadas por meio da ação sequencial de uma DNA glicosilase, uma endonuclease AP, uma desoxirribofosfodiesterase (dRpase), uma DNA polimerase e uma ligase.

Por que as 5-metilcitosinas são *hotspots* para mutações? A desaminação da 5-metilcitosina gera timina (5-metiluracila):



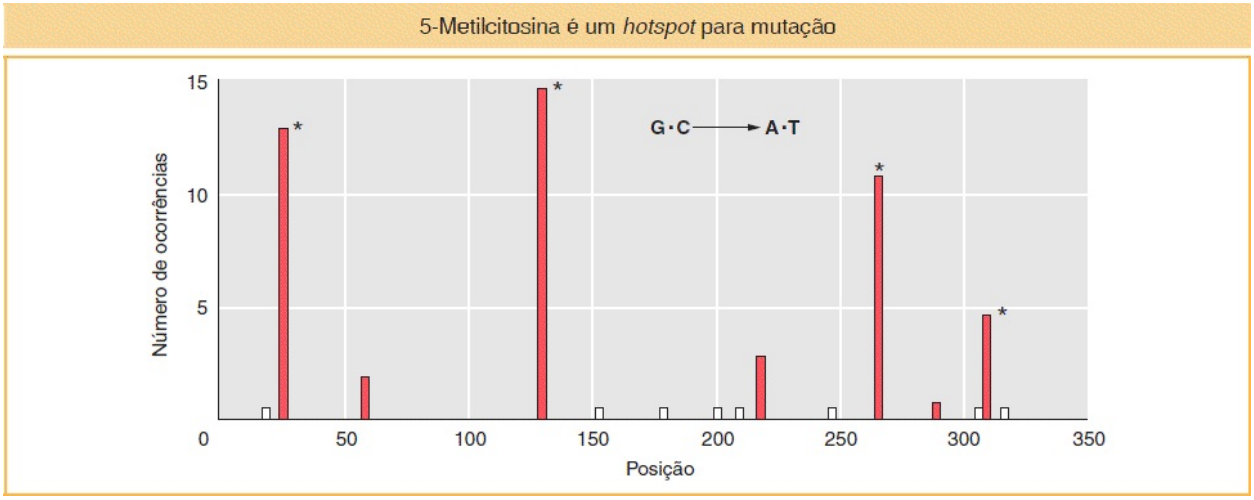


FIGURA 16.21 *Hotspots* de metilcitosina em *E. coli*. Mutações sem sentido em 15 sítios diferentes em *lacI* foram registradas. Todas resultaram em transições G · C → A · T. O *asterisco* (*) marca as posições das 5-metilcitosinas e as *barras brancas* marcam os sítios nos quais transições sabidamente ocorridas não foram isoladas nesse grupo. (Dados de C. Coulondre, J. H. Miller, P. J. Farabaugh e W. Gilbert, "Molecular Basis of Base Substitution Hotspots in *Escherichia coli*", *Nature* 274, 1978, 775.)

A timina não é reconhecida pela enzima uracila-DNA glicosilase e, assim, não é reparada. Portanto, as transições C → T geradas pela desaminação são observadas mais frequentemente nos sítios de 5-metilcitosina em virtude de escaparem a esse sistema de reparo. Uma consequência da frequente mutação da 5-metilcitosina em timina é que as regiões metiladas do genoma (que normalmente

são transcricionalmente inativas; ver Capítulo 12) são convertidas, ao longo do tempo evolutivo, em regiões ricas em AT. Contrariamente, as regiões codificadoras e reguladoras, que são menos metiladas, permanecem ricas em GC.

> CONCEITO-CHAVE No reparo por excisão de base, dano não volumoso ao DNA é reconhecido por uma de diversas enzimas denominadas DNA glicosilases, que clivam as ligações base-açúcar, liberando a base incorreta. O reparo consiste na remoção do sítio, que agora não tem uma base, e na inserção da base correta, conforme orientada pela base complementar no filamento não danificado.

Reparo por excisão de nucleotídio

Embora a maioria dos danos sofridos por um organismo seja um pequeno dano de base, que pode ser reparado por meio do reparo por excisão de base, esse mecanismo não consegue corrigir adições volumosas que distorcem a hélice do DNA, adições tais como os dímeros de pirimidina ciclobutano causados pela luz UV (ver Figura 16.15), nem corrigir dano a mais de uma base. Uma DNA polimerase não consegue continuar a síntese de DNA após tais lesões e, assim, o resultado é um bloqueio na replicação. Uma forquilha de replicação bloqueada pode causar a morte celular. De modo semelhante, uma base anormal ou danificada pode parar o complexo de transcrição. Para lidar com ambas essas situações, procariotos e eucariotos utilizam uma via extremamente versátil, denominada **reparo por excisão de nucleotídio (NER)**, que é capaz de desfazer os bloqueios da replicação e da transcrição e reparar o dano.

Curiosamente, duas doenças autossômicas recessivas em seres humanos, o **xeroderma pigmentoso (XP)** e a **síndrome de Cockayne**, são causadas por defeitos no reparo por excisão de nucleotídio. Embora os pacientes com XP ou síndrome de Cockayne sejam excepcionalmente sensíveis à luz UV, outros sintomas são dramaticamente diferentes. O xeroderma pigmentoso foi introduzido no início deste capítulo e é caracterizado pelo desenvolvimento precoce de

cânceres, especialmente câncer de pele e, em alguns casos, defeitos neurológicos. Contrariamente, os pacientes afetados pela síndrome de Cockayne apresentam uma diversidade de distúrbios do desenvolvimento, incluindo nanismo, surdez e retardo. Em termos amplos, pacientes com XP adquirem câncer precoce, enquanto pacientes com síndrome de Cockayne envelhecem prematuramente. Como defeitos na mesma via de reparo podem levar a sintomas de doença tão diferentes? Embora não exista uma resposta simples para essa questão, os trabalhos sobre a base genética dessas doenças levou à identificação de importantes proteínas na via NER.

O reparo por excisão de nucleotídio é um processo complexo, que requer dúzias de proteínas. Apesar dessa complexidade, o processo de reparo pode ser dividido em quatro fases:

1. Reconhecimento da(s) base(s) danificada(s).
2. Montagem de um complexo multiproteico no sítio.
3. Corte do filamento danificado diversos nucleotídios *upstream* e *downstream* do sítio danificado e remoção dos nucleotídios (aproximadamente 30) entre os cortes.
4. Utilização do filamento não danificado como molde para a DNA polimerase, seguida pela ligação do filamento.

O fato, conforme já mencionado, de a parada tanto da forquilha de replicação como do complexo de transcrição ativar essa via de reparo implica a existência de dois tipos de reparo por excisão de nucleotídio que diferem no reconhecimento do dano (etapa 1). Atualmente sabemos que um tipo, denominado **reparo por excisão de nucleotídio genômico global (GG-NER)**, corrige lesões em qualquer local do genoma e é ativado por forquilhas de replicação paradas. O outro tipo repara as regiões transcritas do DNA e é denominado, não surpreendentemente, **reparo por excisão de nucleotídio acoplado à transcrição (TC-NER)**. Conforme pode ser observado na Figura 16.22, embora a etapa de reconhecimento seja diferente, tanto GG-NER quanto TC-NER compartilham as últimas quatro etapas.

Nesse ponto, você pode estar se perguntando: “E se as diferenças nos sintomas

das doenças XP e síndrome de Cockayne ocorressem em virtude de mutações em diferentes classes de proteínas de reconhecimento?" Você estaria no caminho certo ao indagar essa questão. Pacientes com XP estão contidos em oito grupos de complementação, que carreiam mutações em um dos oito genes que codificam as proteínas XPA a XPG (Figura 16.22). Pacientes com síndrome de Cockayne apresentam uma mutação em uma de duas proteínas denominadas CSA e CSB, que se acredita reconhecerem complexos de transcrição parados.

GG-NER é iniciado quando um complexo proteico de XPC e RAD23B reconhece uma dupla-hélice distorcida causada por uma base danificada e se liga ao filamento oposto. Contrariamente, TC-NER é iniciado quando um complexo de RNA polimerase é parado por uma lesão do DNA no filamento transcrito e CSA e CSB se ligam nesse sítio para formar um complexo de reconhecimento. Após o reconhecimento da lesão, as vias GG-NER e TC-NER utilizam amplamente as mesmas proteínas para remover e reparar o DNA danificado, tendo em vista que o papel de XPC-RAD23B e CSA/CSB é atrair o complexo multiproteico TFIIH. Duas de suas 10 subunidades, XPB e XPD, são helicases (3′ para 5′ e 5′ para 3′, respectivamente), que desenrolam e abrem a hélice de DNA ao redor da lesão. As etapas subsequentes comuns a GG-NER e TC-NER medeiam a clivagem e a excisão da base danificada e até 30 nucleotídios adjacentes, seguidas pela síntese de DNA para preencher o intervalo (ver detalhes na Figura 16.22). Além de XPC-RAD23D, XPB e XPD, pacientes com XP apresentam mutações em outras proteínas que participam nas etapas comuns de NER. Conforme demonstrado na Figura 16.22, XPA promove a liberação da subunidade CAK e a ligação de RPA, enquanto as endonucleases XPF (com ERCCI) e XPG cortam, respectivamente, a 5′ e 3′ do dano ao DNA. Após a remoção da base danificada e do DNA adjacente, o intervalo é preenchido por uma DNA polimerase auxiliada pelas proteínas RFC e PCNA. A última etapa no NER envolve a ligação do novo filamento ao DNA adjacente por meio de um de dois complexos de ligação (XRCC1/LIG3 ou FEN1/LIG1).

As diferenças moleculares entre GG-NER e TC-NER podem proporcionar uma explicação para os diferentes sintomas demonstrados por pacientes com XP e síndrome de Cockayne? Relembre que pacientes com XP desenvolvem cânceres

precoces, enquanto pacientes com síndrome de Cockayne apresentam uma diversidade de sintomas associados ao envelhecimento prematuro. Vimos que o sistema de reparo dos pacientes com síndrome de Cockayne não consegue reconhecer complexos de transcrição parados. Uma consequência desse defeito é que a célula apresenta maior probabilidade de ativar a via suicida da apoptose. Em uma pessoa saudável, a morte celular com frequência é preferível à propagação de uma célula que tenha sofrido dano persistente no DNA. Entretanto, de acordo com essa teoria, a via de morte celular seria ativada mais frequentemente em um paciente com síndrome de Cockayne, levando, assim, a uma diversidade de sintomas de envelhecimento prematuro. Contrariamente, pacientes com XP podem reconhecer os complexos de transcrição parados (eles apresentam proteínas CSA e CSB normais) e evitar a morte celular quando a transcrição é reiniciada. Entretanto, eles não conseguem reparar o dano original em virtude de mutações em uma de suas proteínas XP. Portanto, as mutações serão acumuladas nas células de pacientes com XP e, conforme declarado anteriormente neste capítulo, a presença de mutações, causadas por mutágenos ou por falha das vias de reparo, aumenta o risco de desenvolvimento de muitos tipos de câncer.

> CONCEITO-CHAVE O reparo por excisão de nucleotídio é uma via versátil que reconhece e corrige as lesões do DNA amplamente em virtude dos danos causados por UV e, ao fazer isso, alivia as forquilhas de replicação e os complexos de transcrição parados. Pacientes com xeroderma pigmentoso e síndrome de Cockayne são sensíveis à luz UV em virtude de mutações importantes em proteínas de reparo por excisão de nucleotídio que reconhecem ou reparam as bases danificadas.

Reparo pós-replicação | Reparo de malpareamento

Você aprendeu na primeira metade deste capítulo que ocorrem muitos erros na replicação do DNA. De fato, a taxa de erro é de aproximadamente 10^{-5}. A correção por meio da função de revisão 3′ para 5′ da polimerase replicativa reduz

a taxa de erro para menos de 10^{-7}. A principal via que corrige os erros replicativos remanescentes é denominada **reparo de malpareamento**. Essa via de reparo reduz a taxa de erro para menos de 10^{-9} ao reconhecer e reparar as bases malpareadas e as pequenas alças causadas pela inserção e pela deleção de nucleotídios (indels) no período da replicação. A partir desses valores, você pode verificar que as mutações que levam à perda da via de reparo de malpareamento poderiam aumentar a frequência de mutações em 100 vezes. De fato, a perda do reparo de malpareamento também está associada a formas hereditárias de câncer de cólon.

Os sistemas de reparo de malpareamento devem realizar no mínimo três coisas:

1. Reconhecer pares de bases malpareados.
2. Determinar qual base é a incorreta no malpareamento.
3. Excisar a base incorreta e realizar a síntese de reparo.

Duas vias de reparo por excisão de nucleotídios

Etapa 1: a via do reparo por excisão de nucleotídio (NER) é ativada quando adutos volumosos ou múltiplas bases danificadas são reconhecidos em regiões não transcritas (GG-NER) ou transcritas (TC-NER) do genoma. Na via GG-NER, XPA-RAD23 se liga a uma lesão. Na via TC-NER, CSA e CSB se ligam a complexos de RNA polimerase II que estão parados em uma lesão

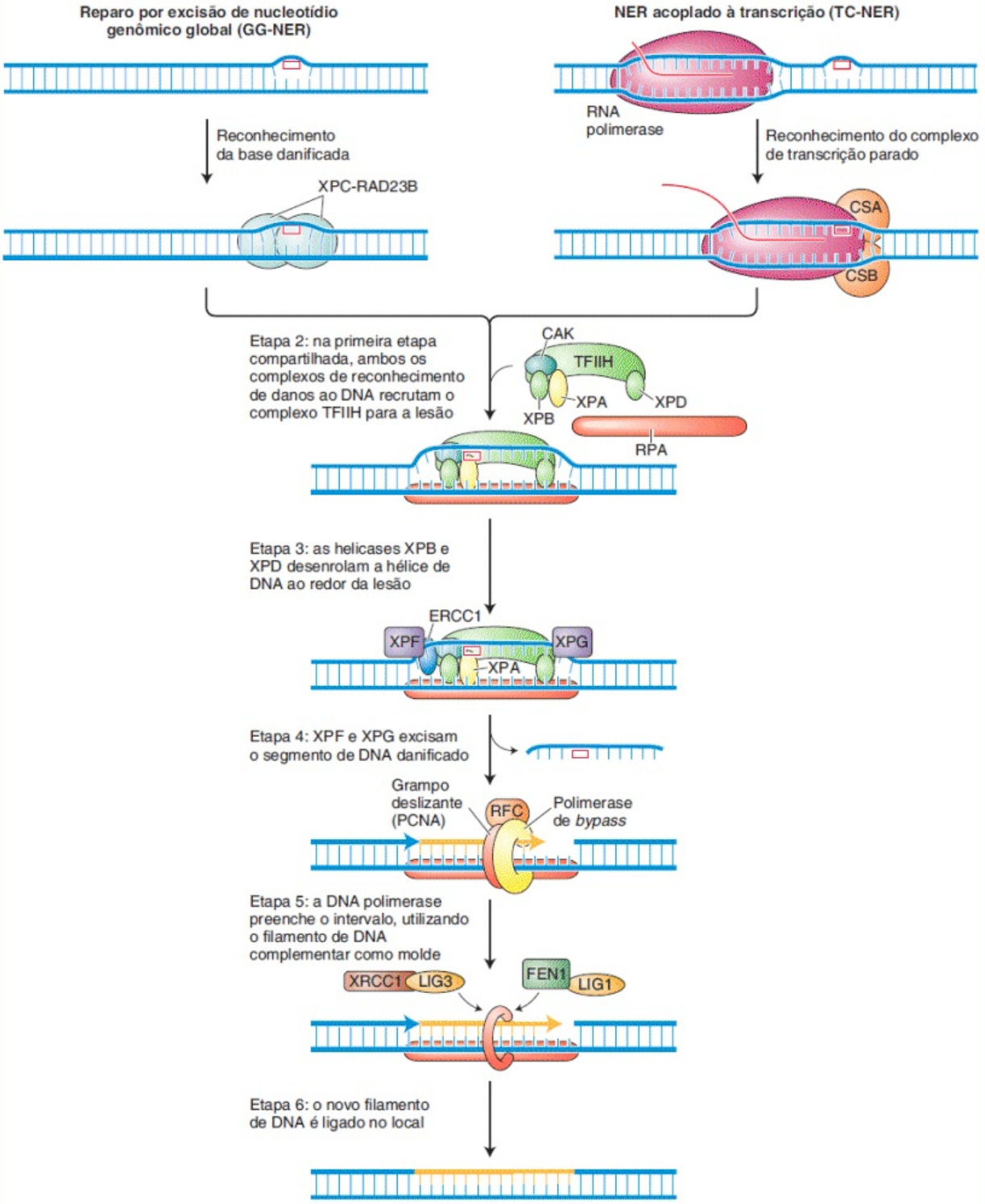


FIGURA 16.22 A via de reparo por excisão de nucleotídio é ativada quando adutos volumosos ou múltiplas bases danificadas são reconhecidos em regiões não transcritas (GG-NER) ou transcritas (TC-NER) do genoma. Essas duas vias são iniciadas por diferentes eventos e complexos distintos, conforme demonstrado na primeira etapa. Ambos os complexos atuam para atrair o mesmo complexo TFIIH (Etapa 2). Os complexos de reconhecimento são removidos na Etapa 3 (não demonstrada). Nas etapas 4 a 6, um complexo multiproteico excisa diversas bases e as ressintetiza utilizando o filamento oposto como molde. Ver texto para detalhes.

A maior parte do que se sabe a respeito do reparo de malpareamento tem origem em décadas de análises genéticas e bioquímicas com a utilização da bactéria-modelo *E. coli* (ver Organismo-modelo, *E. coli*, no Capítulo 5). Especialmente digna de nota foi a reconstituição do sistema de reparo de malpareamento em tubo de ensaio no laboratório de Paul Modrich. A conservação de muitas das proteínas de reparo de malpareamento, de bactérias e leveduras até seres humanos, indica que essa via é antiga e importante em todos os organismos vivos. Recentemente, o sistema de reparo de malpareamento humano também foi reconstituído em tubo de ensaio no laboratório de Modrich. A capacidade de estudar os detalhes da reação estimulará futuros estudos sobre a via humana. Entretanto, por enquanto enfocaremos o sistema muito bem-caracterizado de *E. coli* (Figura 16.23).

A primeira etapa no reparo de malpareamento é o reconhecimento do dano no DNA recém-replicado pela proteína MutS. A ligação dessa proteína às distorções na dupla-hélice de DNA causadas por bases incompatíveis inicia a via de reparo de malpareamento ao atrair três outras proteínas para o sítio da lesão (MutL, MutH e UvrD [não demonstrada]). A proteína-chave é MutH, que realiza a função crucial de cortar o filamento que contém a base incorreta. Sem essa capacidade de discriminação entre as bases corretas e incorretas, o sistema de reparo de malpareamento não poderia determinar qual base deve ser excisada para evitar o surgimento de uma mutação. Se, por exemplo, ocorre malpareamento entre G-T como um erro de replicação, como o sistema consegue determinar se G ou T está incorreta? Ambas são bases normais no DNA. Mas os erros de replicação produzem pareamentos errados no filamento recém-sintetizado e, assim, o sistema de reparo de malpareamento substitui a base naquele filamento.

Como o reparo de malpareamento distingue o filamento recém-sintetizado do antigo? Relembre do Capítulo 12 que as bases citosina com frequência são

metiladas em eucariotos e que essa marca epigenética é propagada do filamento parental para o filho logo após a replicação. O DNA de *E. coli* também é metilado, mas os grupos metil relevantes para o reparo de malpareamento são adicionados às bases adenina. Para distinguir o filamento-molde antigo do filamento recém-sintetizado, o sistema de reparo bacteriano se aproveita de um atraso na metilação da sequência a seguir:

5′-G-A-T-C-3′
3′-C-T-A-G-5′

A enzima responsável pela metilação é a adenina metilase, que cria a 6-metiladenina em cada filamento. Entretanto, a adenina metilase necessita de diversos minutos para reconhecer e modificar os trechos de GATC recém-sintetizados. Nesse intervalo, a proteína MutH corta o sítio de metilação no filamento que contém A ainda não metilada. Esse sítio pode estar a diversas centenas de pares de bases de distância da base erroneamente pareada. Após o corte no sítio, a proteína UrvD se liga ao corte e utiliza a sua atividade de helicase para desenrolar o DNA. Uma proteína de ligação a filamento único protetora reveste o filamento parental desenrolado, enquanto a parte do novo filamento entre o pareamento errado e o corte é excisada.

O reparo de malpareamento corrige os erros replicativos

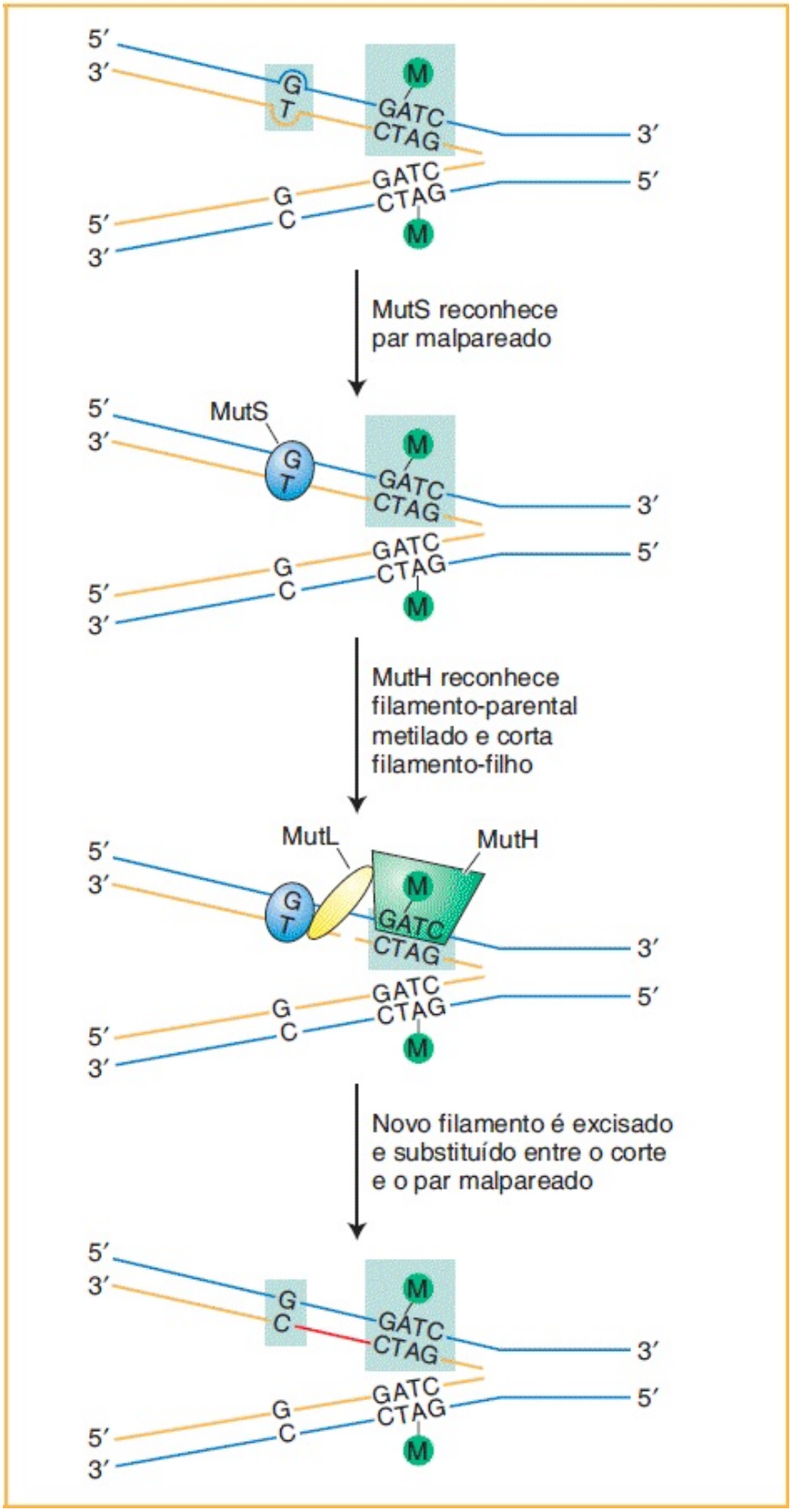


FIGURA 16.23 Modelo em relação ao reparo de malpareamento em *E. coli*. O DNA é metilado (Me) no resíduo A na sequência GATC. A replicação do DNA produz um dúplex hemimetilado que existe até que a metilase consiga modificar o filamento recém-sintetizado. O sistema de reparo de malpareamento realiza quaisquer correções necessárias com base na sequência observada no filamento metilado (modelo original). MutS, MutH e MutL são proteínas.

Muitas das proteínas no reparo de malpareamento de *E. coli* são conservadas no reparo de malpareamento humano. No entanto, como os eucariotos reconhecem e reparam apenas o filamento recém-sintetizado ainda é desconhecido. O problema causa perplexidade nos organismos sem a maior parte da metilação do DNA, ou completamente desprovidos dela, tais como leveduras, *Drosophila* e *C. elegans*. Um modelo popular propõe que a discriminação tenha por base o reconhecimento de extremidades livres 3′ que caracterizam os filamentos de replicação contínua e descontínua recém-sintetizados.

Um alvo importante do sistema de malpareamento humano são as curtas sequências repetidas, que podem ser expandidas ou deletadas na replicação pelo mecanismo de malpareamento descrito anteriormente (ver Figura 16.8). Mutações em alguns dos componentes dessa via demonstraram ser responsáveis por diversas doenças humanas, especialmente cânceres. Existem milhares de repetições curtas (microssatélites) localizadas por todo o genoma humano (ver Capítulo 4). Embora a maior parte delas esteja localizada em regiões não codificadoras (tendo em vista que a maior parte do genoma é não codificador), algumas estão localizadas em genes críticos para o crescimento e o desenvolvimento normais.

Portanto, pode-se prever que defeitos na via de reparo de malpareamento humana tenham como consequência doenças muito graves. Essa previsão comprovou ser verdadeira, e um caso pontual é a síndrome denominada câncer colorretal não polipose hereditário (HNPCC), a qual, apesar de sua denominação, não é propriamente um câncer, mas aumenta o risco de câncer. Uma das predisposições hereditárias ao câncer mais comuns, a doença afeta uma em 200 pessoas no mundo ocidental. Estudos demonstraram que o HNPCC resulta da perda do sistema de reparo de malpareamento, em grande parte em virtude de

mutações herdadas em genes que codificam as contrapartes (e homólogas) humanas das proteínas MutS e MutL bacterianas (ver Figura 16.23). A herança do HNPCC é autossômica dominante. As células com uma cópia funcional dos genes de reparo de malpareamento apresentam atividade de reparo de malpareamento normal, mas as linhagens celulares tumorais têm origem em células que perderam uma cópia funcional e que, portanto, são deficientes do malpareamento. Essas células demonstram altas taxas de mutação, em parte em virtude de uma incapacidade de corrigir a formação de indel na replicação.

> CONCEITO-CHAVE O sistema de reparo de malpareamento corrige erros na replicação que não são corrigidos pela função de revisão da DNA polimerase replicativa. O reparo é restrito ao filamento recém-sintetizado, que é reconhecido pelo maquinário de reparo em procariotos, porque não tem um marcador de metilação.

Reparo propenso a erro | Síntese de DNA translesão

Até agora, todos os mecanismos de reparo que encontramos são livres de erro, uma vez que eles revertem o dano diretamente, ou utilizam a complementaridade de bases para inserir a base correta. Além disso, existem vias de reparo que são, elas próprias, uma fonte significativa de mutação. Esses mecanismos aparentam ter evoluído para evitar a ocorrência de desfechos possivelmente mais sérios, tais como morte celular ou câncer. Conforme já mencionado, uma forquilha de replicação parada pode iniciar uma via de morte celular. Tanto em procariotos quanto em eucariotos, tais blocos de replicação podem ser *contornados* pela inserção de bases inespecíficas. Em *E. coli*, esse processo requer a ativação do **sistema SOS**. A denominação *SOS* tem origem na ideia de que esse sistema é induzido como uma resposta de emergência para prevenir a morte celular na presença de dano significativo ao DNA. Como tal, a indução de SOS é um mecanismo de último recurso, um tipo de tolerância ao dano, que possibilita que uma célula troque a morte por um determinado nível de mutagênese.

Levou mais de 30 anos para descobrir como o sistema SOS dá origem a mutações enquanto possibilita que a DNA polimerase contorne lesões nas forquilhas de replicação paradas. Já estamos familiarizados com o dano do DNA induzido pela luz UV (ver Figura 16.15). Uma classe incomum de mutantes de *E. coli* que sobreviveram à exposição à luz UV sem a sofrer mutações adicionais foi isolada na década de 1970. O fato de que tais mutantes existiram sugeriu que alguns genes de *E. coli* atuem para gerar mutações quando expostos à luz UV. A mutação induzida pela luz UV não ocorrerá se os genes *DinB, UmuC,* ou *UmuD'* estiverem mutados.

A Figura 16.24 demonstra as etapas no mecanismo SOS. Na primeira etapa, a luz UV induz a síntese de uma proteína denominada RecA. Veremos mais sobre a proteína RecA posteriormente no capítulo, tendo em vista que ela é um componente importante nos mecanismos-chave de reparo do DNA e recombinação. Quando a polimerase replicativa (DNA polimerase III) está parada em um sítio de dano ao DNA, o DNA à frente da polimerase continua a ser desenrolado, expondo regiões de DNA unifilamentares que se ligam a proteínas de ligação a filamento único. Em seguida, as proteínas RecA unem-se às proteínas de ligação a filamento único e formam um filamento de proteína–DNA. O filamento RecA é a forma biologicamente ativa dessa proteína. Nessa situação, RecA atua como um sinal que leva à indução de diversos genes sabidamente reconhecidos por codificar membros de uma família recém-descoberta de DNA polimerases, que conseguem contornar o bloqueio da replicação e que são distintas das polimerases replicativas. As DNA polimerases que conseguem contornar as paradas da replicação também foram observadas em diversos grupos de eucariotos, que variam de leveduras a seres humanos. Essas polimerases eucarióticas contribuem para um mecanismo de tolerância a danos, denominado **síntese de DNA translesão**, que se assemelha ao sistema de *bypass* SOS em *E. coli*.

Essas **polimerases translesão**, ou ***de bypass***, como passaram a ser conhecidas, diferem das principais polimerases replicativas de diversos modos. Primeiramente, elas conseguem tolerar adutos incomumente grandes nas bases. Enquanto a polimerase replicativa para se uma base não se ajusta no sítio ativo,

as polimerases de *bypass* apresentam bolsões muito maiores, que conseguem acomodar bases danificadas. Em segundo lugar, em algumas situações, as polimerases de *bypass* apresentam uma taxa de erro muito mais alta, em parte em virtude da ausência da atividade de revisão 3′ para 5′ das principais polimerases replicativas. Por fim, elas conseguem adicionar apenas alguns nucleotídios antes de sair. Essa característica é atraente, tendo em vista que a principal função de uma polimerase propensa a erro é desbloquear a forquilha de replicação, não sintetizar longos trechos de DNA que poderiam conter muitos pareamentos errados.

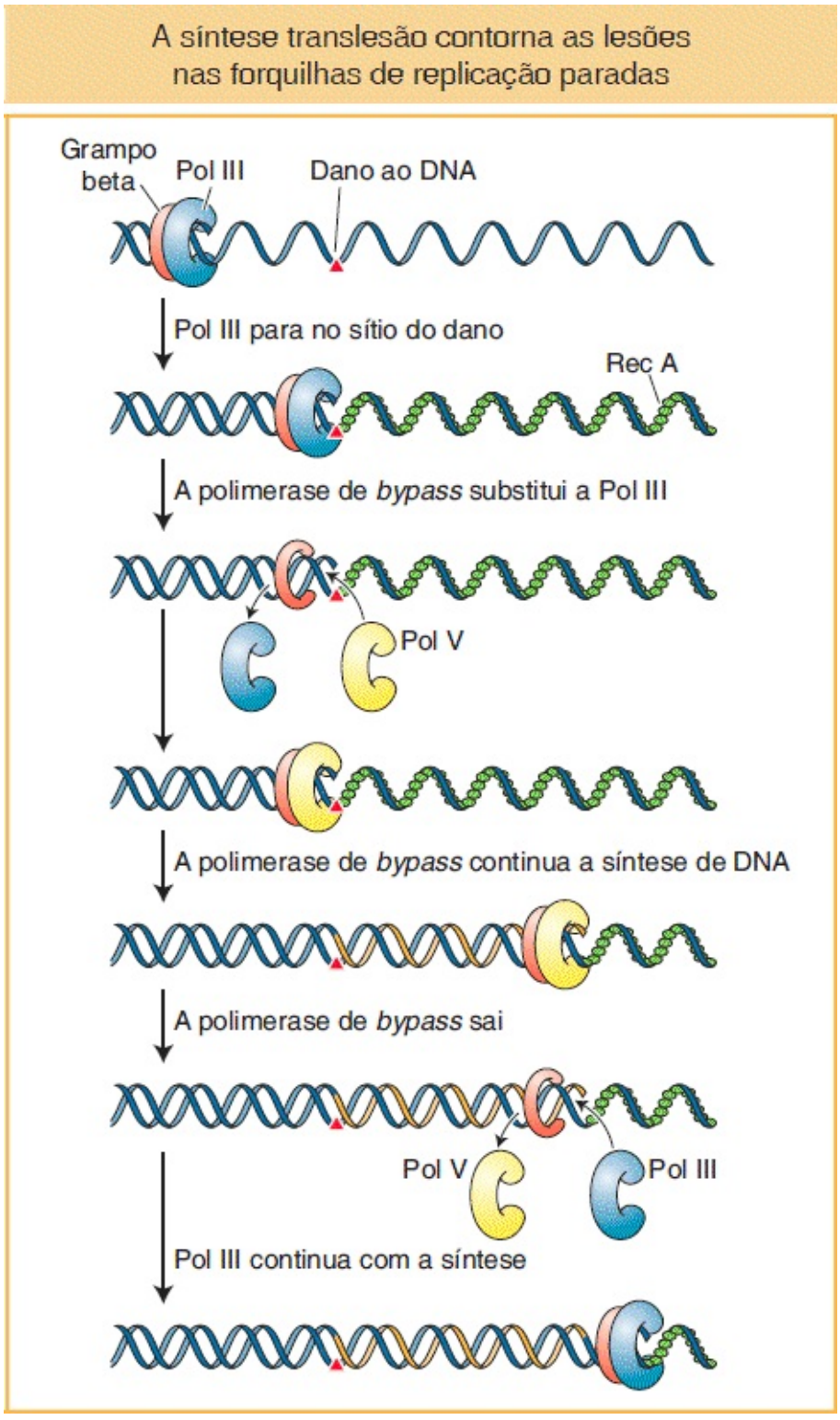


FIGURA 16.24 Um modelo de síntese translesão em *E. coli*. Durante o período da replicação, a DNA polimerase III é temporariamente substituída por uma polimerase de *bypass* (pol V), que consegue continuar a replicação, passando pela lesão. As polimerases de *bypass* são propensas a erros. O grampo β bacteriano

(*proteína vermelha*) é equivalente à PCNA eucariótica. (Dados de E. C. Friedberg, A. R. Lehmann e R. P. Fuchs, "Trading Places: How Do DNA Polymerases Switch During Translesion DNA Synthesis?" *Molec. Cell* 18, 2005, 499-505.)

Atualmente são conhecidas diversas polimerases de *bypass* que aparentam estar sempre presentes em células eucarióticas. Em virtude de estarem sempre presentes, seu acesso ao DNA deve ser regulado, de modo que elas sejam utilizadas apenas quando necessário. A célula desenvolveu uma solução organizada para esse problema. Relembre que uma parte integrante do replissomo é a proteína PCNA (antígeno nuclear de proliferação celular), que atua como um grampo deslizante para orquestrar os incontáveis eventos na forquilha de replicação (ver Figura 7.20). Uma proteína crítica presente em uma forquilha de replicação parada é a Rad6, a qual, curiosamente, é uma enzima que adiciona ubiquitina às proteínas (Figura 16.25). Conforme descrito no Capítulo 9, a adição de cadeias de muitos monômeros de ubiquitina marca a proteína para degradação (ver Figura 9.23). Contrariamente, a ligação de um monômero de ubiquitina único à PCNA altera sua conformação, de modo que agora ela consegue se ligar à polimerase de *bypass* e orquestrar a síntese translesão. A remoção enzimática da marcação de ubiquitina na PCNA leva à dissociação da polimerase de *bypass* e à restauração final da replicação normal. Qualquer malpareamento de base em virtude da síntese translesão ainda apresenta uma chance de detecção e correção pela via de reparo de malpareamento.

A regulação da função de PCNA por meio da adição e da remoção de monômeros de ubiquitina ilustra a importância das modificações pós-traducionais em eucariotos. Se o dano da base no filamento-molde não for rapidamente corrigido, a forquilha de replicação parada sinalizará a ativação da via de morte celular. Uma célula eucariótica não pode esperar pela síntese *de novo* de polimerases de *bypass* após a transcrição e a tradução, como ocorre no sistema SOS de *E. coli*. Em vez disso, as polimerases de *bypass* eucarióticas são transcritas constitutivamente e estão sempre presentes; o seu acesso à forquilha de replicação é controlado por modificações pós-traducionais rápidas e reversíveis.

CONCEITO-CHAVE Na síntese translesão, polimerases de *bypass* são recrutadas para as forquilhas de replicação que pararam em virtude do dano no filamento-molde. As polimerases de *bypass* introduzem erros no período da síntese, que podem persistir e levar à mutação, ou que podem ser corrigidos por meio de outros mecanismos, tais como o reparo de malpareamento.

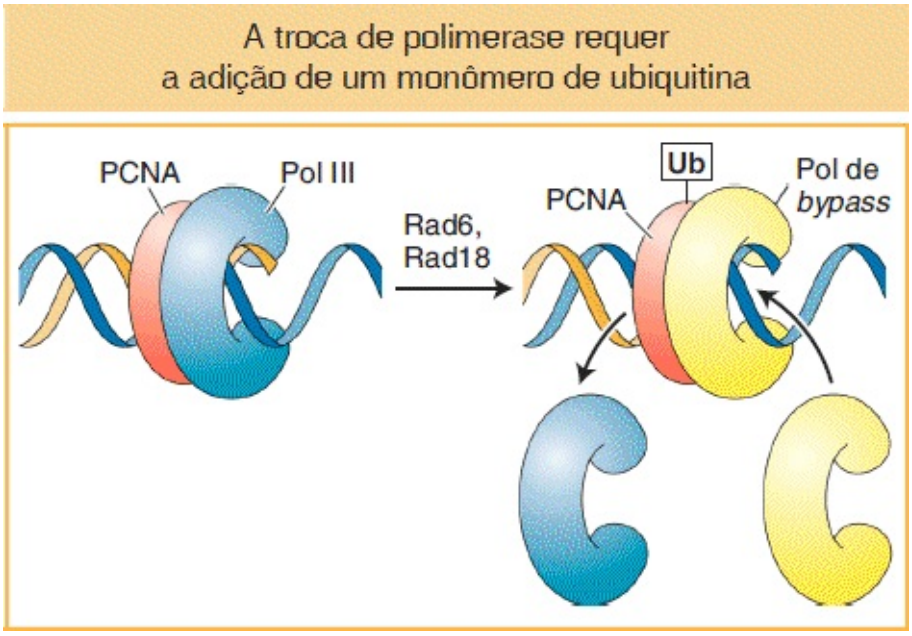


FIGURA 16.25 A adição de um único monômero de ubiquitina (Ub) ao grampo deslizante (PCNA) possibilita que a polimerase de *bypass* se ligue ao PCNA e inicie a replicação.

Reparo de quebras bifilamentares

Conforme vimos, muitos sistemas de correção exploram a complementaridade do DNA para realizar reparos livres de erros. Tal reparo livre de erros é caracterizado por dois estágios: (1) remoção das bases danificadas, talvez em conjunto com o DNA próximo, de um filamento da dupla-hélice e (2) utilização do outro filamento como um molde para a síntese de DNA necessária para preencher o espaço unifilamentar. Entretanto, o que ocorreria se *ambos* os filamentos da dupla-hélice estivessem danificados de tal modo que a complementaridade não pudesse ser explorada? Por exemplo, a exposição aos

raios X com frequência causa a quebra de ambos os filamentos da dupla-hélice em sítios bem próximos. Esse tipo de mutação é denominado **quebra bifilamentar**. Se não reparadas, as quebras bifilamentares podem causar uma diversidade de aberrações cromossômicas, resultando em morte celular ou um estado pré-canceroso.

Curiosamente, a produção de quebras bifilamentares é uma característica integrante de alguns processos celulares normais que necessitam de rearranjos de DNA. Um exemplo é a recombinação meiótica. Conforme veremos no restante deste capítulo, a célula utiliza muitas das mesmas proteínas e vias para reparar as quebras bifilamentares e para realizar a recombinação meiótica. Por esse motivo, iniciamos enfocando os mecanismos moleculares que reparam as quebras bifilamentares antes de voltarmos a nossa atenção para o mecanismo de recombinação meiótica.

As quebras bifilamentares podem surgir espontaneamente (p. ex., em resposta a espécies reativas de oxigênio produzidas como subproduto do metabolismo celular), ou podem ser induzidas por radiação ionizante. São conhecidos diversos mecanismos para o reparo de quebras bifilamentares e novos mecanismos ainda estão sendo descobertos. Dois mecanismos distintos estão descritos na seção a seguir: junção de extremidades não homólogas e recombinação homóloga.

Junção de extremidades não homólogas. Muitos dos mecanismos de reparo anteriormente descritos são ativados na fase S do ciclo celular, quando o DNA está se replicando em preparo para a mitose ou a meiose. Entretanto, contrariamente às células da maior parte dos procariotos e dos eucariotos inferiores, as células dos eucariotos superiores normalmente não estão replicando seu DNA, tendo em vista que estão em uma fase de repouso do ciclo celular, ou cessaram totalmente de se dividir. O que acontece quando ocorrem quebras bifilamentares em células nas quais os filamentos ou as cromátides-irmãs não danificados não estão presentes? A resposta é que essas extremidades devem ser reparadas, perfeita ou imperfeitamente, tendo em vista que as extremidades quebradas podem iniciar rearranjos cromossômicos possivelmente perigosos, que podem levar a um estado canceroso (ver Capítulo 17).

Um modo pelo qual os eucariotos superiores juntam as extremidades bifilamentares quebradas é um mecanismo um tanto deselegante, porém importante, denominado **junção de extremidades não homólogas (NHEJ)**, que está demonstrado na Figura 16.26. Assim como em outros mecanismos de reparo, a primeira etapa na via NHEJ é o reconhecimento do dano. A via NHEJ é iniciada quando duas proteínas muito abundantes, KU70 e KU80, ligam-se às extremidades quebradas, formando um heterodímero que atua em duas funções. Primeiramente, ele previne lesão adicional às extremidades e, em segundo lugar, recruta outras proteínas (em verde, na Figura 16.26) que cortam as extremidades do filamento para gerar as extremidades 5′-P e 3′-OH que são necessárias para a ligação. A DNA ligase IV em seguida une as duas extremidades.

Como os cientistas sabem quando todos os componentes de uma via biológica foram identificados? Como se vê, esse problema é difícil. A recente identificação de um novo componente da via NHEJ fornece um exemplo.

Por diversos motivos, acreditava-se que todos os componentes da via NHEJ tivessem sido identificados. Entretanto, geneticistas que analisavam uma linhagem celular (denominada 2BN) derivada de uma criança com um distúrbio hereditário raro foram surpreendidos. Embora eles pudessem demonstrar que aquela linhagem celular 2B era defeituosa em relação ao reparo de quebras bifilamentares, eles não conseguiram restaurar o sistema de reparo e produzir o fenótipo do tipo selvagem por meio de complementação genética com quaisquer dos genes que codificam as proteínas de NHEJ. Ou seja, quando eles introduziram genes do tipo selvagem que codificam proteínas NHEJ conhecidas (p. ex., KU70, KU80, ligase IV) na linhagem 2BN, a linhagem celular ainda era defeituosa no reparo das quebras bifilamentares. Esse resultado negativo indicou que a linhagem celular 2BN carreava uma mutação em uma proteína NHEJ desconhecida.

Na era da genômica, a identificação de proteínas ligadas a doenças está se tornando mais comum em virtude da ampla disponibilidade de linhagens celulares de pessoas com fenótipos de doenças. Quando nós, seres humanos, apresentamos um problema de saúde, vamos a um médico e lhe contamos a respeito dos nossos sintomas, incluindo as informações a respeito de parentes com problemas

semelhantes. Tais informações são de crescente importância na era da genômica, com sua *toolkit* genética sempre em expansão, que com frequência pode ser utilizada para identificar genes mutantes associados a distúrbios hereditários (ver Capítulos 10 e 14).

O QUE OS GENETICISTAS FAZEM ATUALMENTE

O que os geneticistas podem fazer em laboratório para encontrar uma proteína, tal como a proteína NHEJ desconhecida, que ainda não tenha sido identificada? Dois laboratórios que utilizam abordagens muito diferentes obtiveram sucesso na identificação do novo componente NHEJ; uma abordagem será descrita aqui, tendo em vista que ela tem sido empregada com sucesso para descobrir diversas outras proteínas. Conforme observado nos capítulos precedentes, muitas proteínas celulares realizam suas ações por meio da interação com outras proteínas. O Capítulo 14, por exemplo, descreveu o teste di-híbrido de levedura, utilizado para identificar proteínas que interagem com a proteína de interesse. No caso atualmente sob consideração, a proteína de interesse era o componente XRCC4 da NHEJ (ver Figura 16.26) e o teste di-híbrido identificou uma proteína de interação 33-kD que era codificada por uma matriz de leitura aberta humana não caracterizada. O fato de duas proteínas interagirem no teste di-híbrido de levedura não significa necessariamente que essas proteínas interagem nas células humanas. Para estabelecer uma conexão entre a proteína 33-kD e a via NHEJ, os geneticistas utilizaram outra técnica valiosa da sua caixa de ferramentas, a do RNAi (ver Capítulo 8). Nesse caso, eles demonstraram que as células normais que expressam RNA antissenso da ORF que codifica a proteína 33-kD, que evitaria a tradução desse gene em proteína, agora eram defeituosas na execução da via NHEJ.

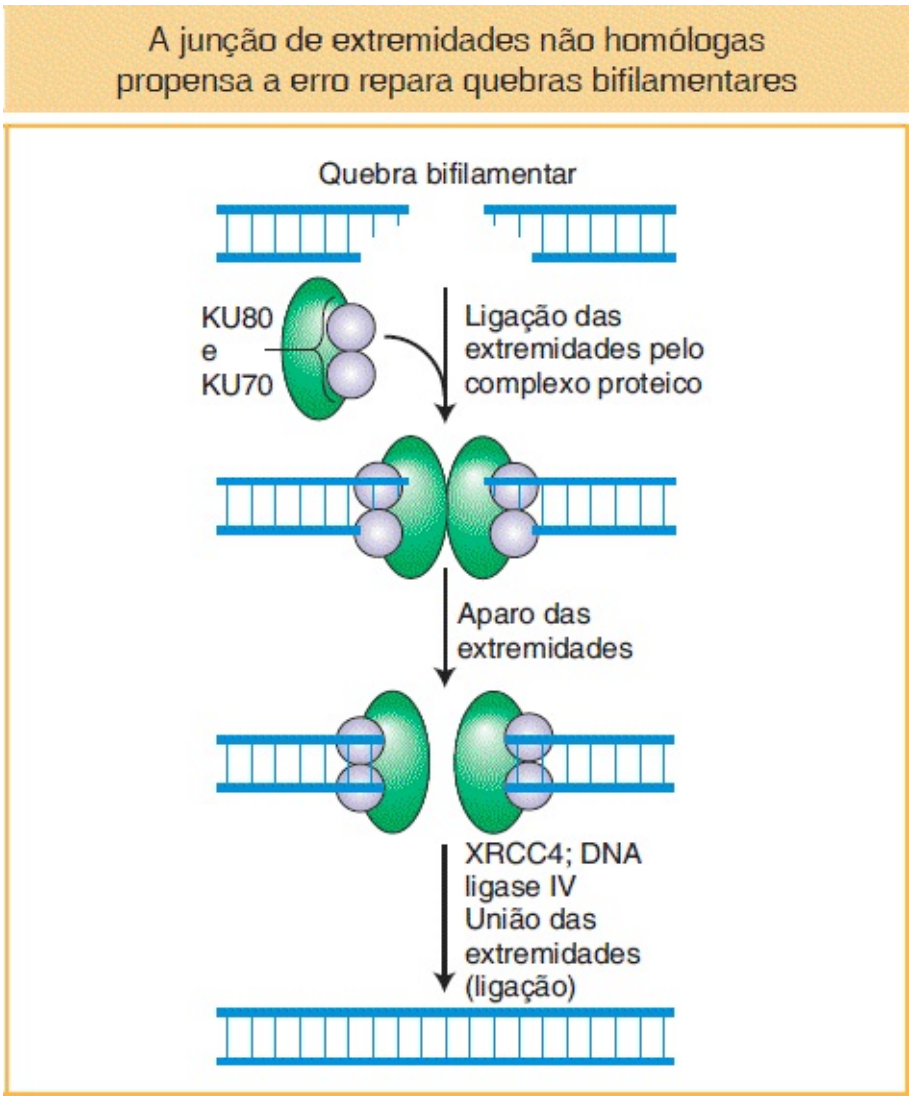


FIGURA 16.26 Mecanismo de junção de extremidades não homólogas (NHEJ). Esse mecanismo é propenso a erro. Ver texto para detalhes.

Nessa história, o círculo foi fechado quando foi demonstrado que as células 2BN defeituosas no reparo bifilamentar apresentam ausência da proteína 33-kD. A expressão dessa proteína corrigiu os defeitos celulares.

> CONCEITO-CHAVE NHEJ é uma via propensa a erro que repara quebras bifilamentares em eucariotos superiores ligando as pontas livres de volta. A identificação dos genes responsáveis por distúrbios hereditários é uma via importante utilizada pelos geneticistas para isolar componentes anteriormente

> desconhecidos das vias de reparo e de outras vias biológicas.

Recombinação homóloga. Se uma quebra bifilamentar ocorre após a replicação de uma região cromossômica em uma célula em divisão, o dano pode ser corrigido por meio de um mecanismo livre de erros denominado **pareamento de filamento dependente de síntese (SDSA)**. Esse mecanismo está ilustrado na Figura 16.27. Ele utiliza as cromátides-irmãs disponíveis na mitose como moldes para assegurar o reparo correto.

As primeiras etapas no SDSA são a ligação das extremidades quebradas por meio de proteínas especializadas e enzimas, o aparo das extremidades 5′ por uma endonuclease para expor as regiões unifilamentares e o revestimento dessas regiões com proteínas que incluem o homólogo de RecA, Rad51. Relembre que na resposta SOS, os monômeros RecA se associam a regiões de DNA unifilamentar para formar filamentos de nucleoproteína. De modo semelhante, Rad51 forma longos filamentos na medida em que se associa à região unifilamentar exposta. Em seguida o filamento de DNA-Rad51 participa em uma extraordinária busca pela cromátide-irmã não danificada para a sequência complementar que será utilizada como molde para a síntese de DNA. Esse processo é denominado invasão de filamento. A extremidade 3′ do filamento invasor desloca uma das cromátides-irmãs não danificada, que forma uma alça D (em referência ao deslocamento) e inicia a síntese de DNA a partir de sua extremidade 3′ livre. A síntese do novo DNA continua a partir de ambas as extremidades 3′ até que ambos os filamentos se desenrolem de seus moldes e se pareiem. A ligação sela os cortes, deixando um trecho reparado de DNA que apresenta uma característica muito distinta: ele foi replicado por meio de um processo conservativo. Ou seja, ambos os filamentos foram recém-sintetizados, o que está em acentuado contraste com a replicação semiconservativa da maior parte do DNA (ver Capítulo 7).

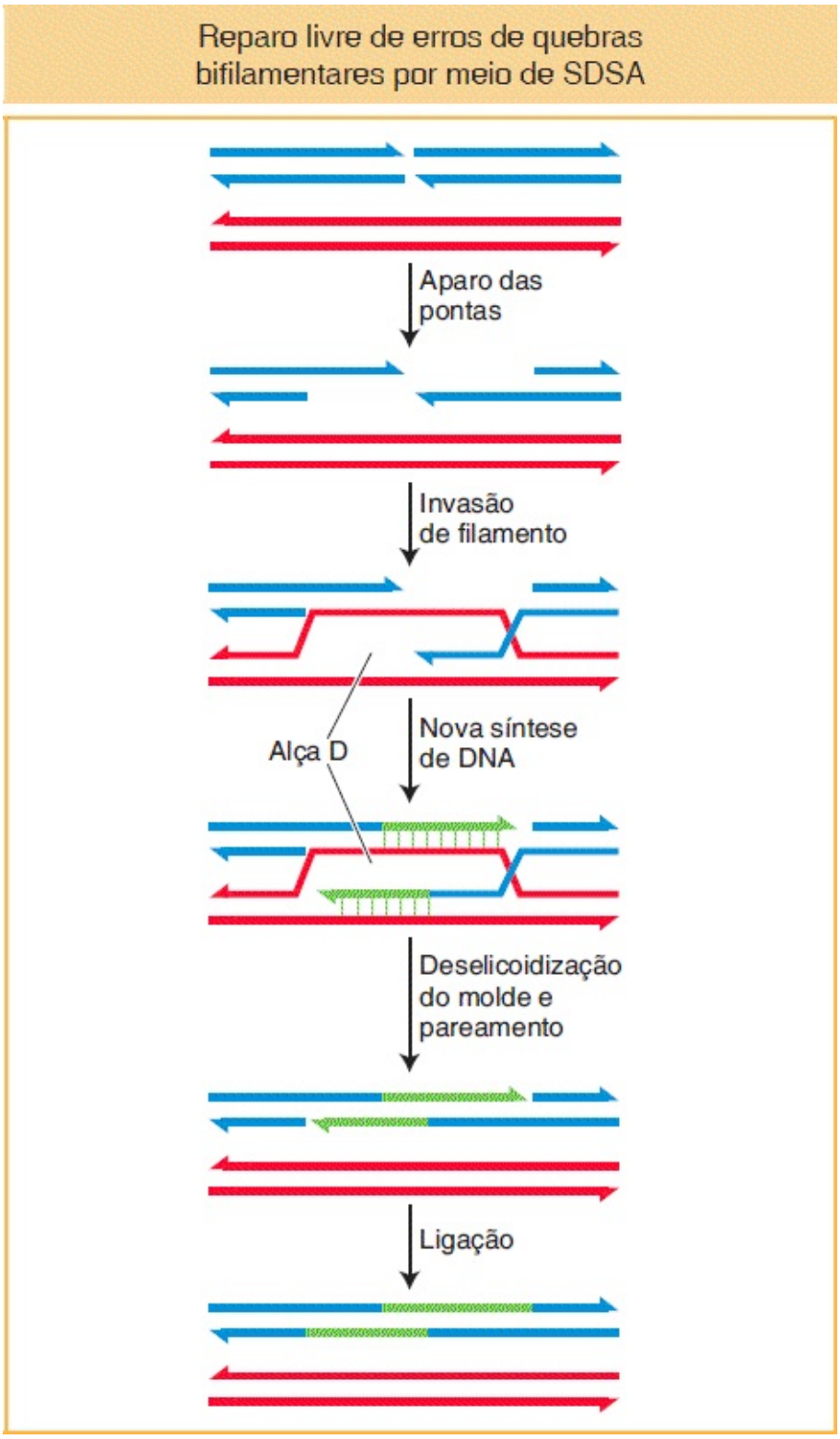


FIGURA 16.27 O mecanismo sem erros de pareamento de filamento dependente de síntese (SDSA) repara quebras bifilamentares em células em divisão.

> CONCEITO-CHAVE O pareamento de filamento dependente de síntese é um mecanismo livre de erros que repara quebras bifilamentares em células em divisão nas quais uma cromátide-irmã está disponível para atuar como molde para a síntese de reparo.

Envolvimento do reparo DSB na recombinação meiótica

Nossa consideração sobre o reparo de quebras bifilamentares em células em divisão leva naturalmente ao tópico do *crossing over* na meiose, tendo em vista que uma quebra bifilamentar inicia o *crossing over*. Embora as quebras sejam uma parte normal e essencial da meiose, elas são, se não processadas correta e eficientemente, tão perigosas quanto as quebras acidentais discutidas até agora.

O *crossing over* é um processo notavelmente preciso que ocorre entre dois cromossomos homólogos (Figura 16.28). Tal processo foi descrito na Seção 4.8. Relembre que a recombinação ocorre após a forquilha de replicação ter passado por uma região cromossômica, formando duas cromátides de cada cromossomo homólogo. Uma cromátide de um cromossomo homólogo irá recombinar com uma cromátide não irmã do outro cromossomo homólogo. Para que a segregação meiótica atue corretamente, cada par de homólogos deve apresentar no mínimo um *crossing over*. A recombinação é iniciada quando uma enzima denominada Spo11 realiza cortes no DNA bifilamentar em uma das cromátides que irá recombinar (Figura 16.29). Embora tenha sido descoberta pela primeira vez em leveduras, a proteína Spo11 é amplamente conservada em eucariotos, indicando que esse mecanismo para iniciar a recombinação é amplamente empregado.

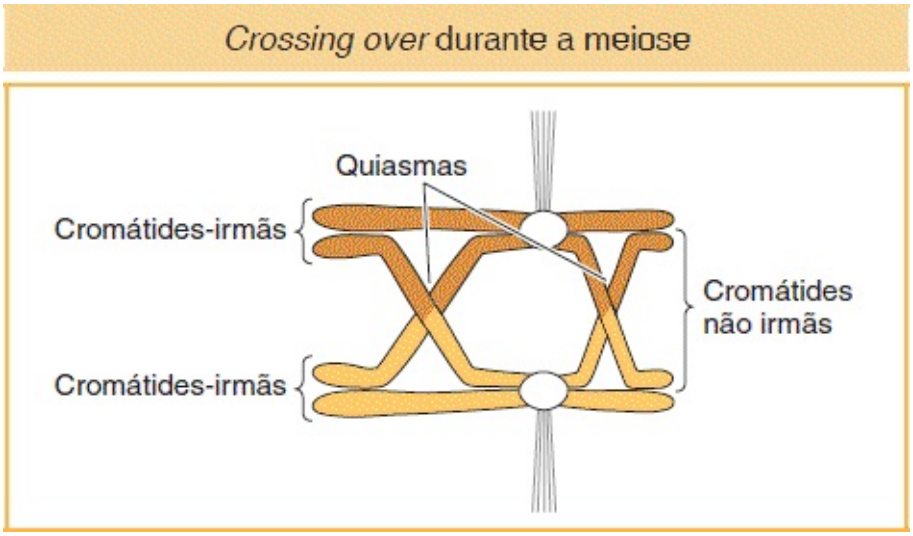


FIGURA 16.28 A troca de braços cromossômicos entre cromátides não irmãs durante a meiose produz um quiasma, o local dos *crossing overs*. Os círculos representam os centrômeros que estão unidos às fibras do fuso.

Após a realização de seus cortes, a enzima Spo11 permanece unida à extremidade 5′, agora livre, onde aparentemente atua para duas finalidades. Primeiramente, ela protege as extremidades contra danos adicionais, incluindo a recombinação espúria com outras extremidades livres. Em segundo lugar, ela pode atrair outras proteínas que são necessárias para a próxima etapa na recombinação. Essa etapa de fato é muito semelhante ao que ocorre no reparo de quebras bifilamentares nas células em divisão. As extremidades 5′ são aparadas novamente (resseccionadas) e um complexo proteico se liga às extremidades 3′ unifilamentares (ver Figura 16.29). Esse complexo inclui a proteína Rad51, a qual, conforme já mencionado, é homóloga à proteína RecA que participa nessa extraordinária busca pela complementaridade na cromátide-irmã.

Nesse momento, a recombinação meiótica assume uma via dramaticamente diferente do reparo de quebra bifilamentar. Na meiose, Rad51 se associa a outra proteína, Dmc1, que está presente apenas durante a meiose (ver Figura 16.29). (Deve-se observar que os organismos-modelo *Drosophila* e *C. elegans* não apresentam homólogos da Dmc1). De alguma maneira, por meio de um mecanismo não completamente compreendido, o filamento que contém Rad51 e Dmc1 conduz uma busca por uma sequência complementar. Entretanto,

contrariamente ao reparo de quebra bifilamentar, o filamento busca uma cromátide não irmã do cromossomo homólogo e não a cromátide-irmã. A busca culmina com a invasão do filamento e a formação da alça D, assim como no reparo de quebra bifilamentar. Esses eventos são necessários para a formação de quiasmas na meiose I. Ou seja, os homólogos se tornam conectados como um resultado da recombinação.

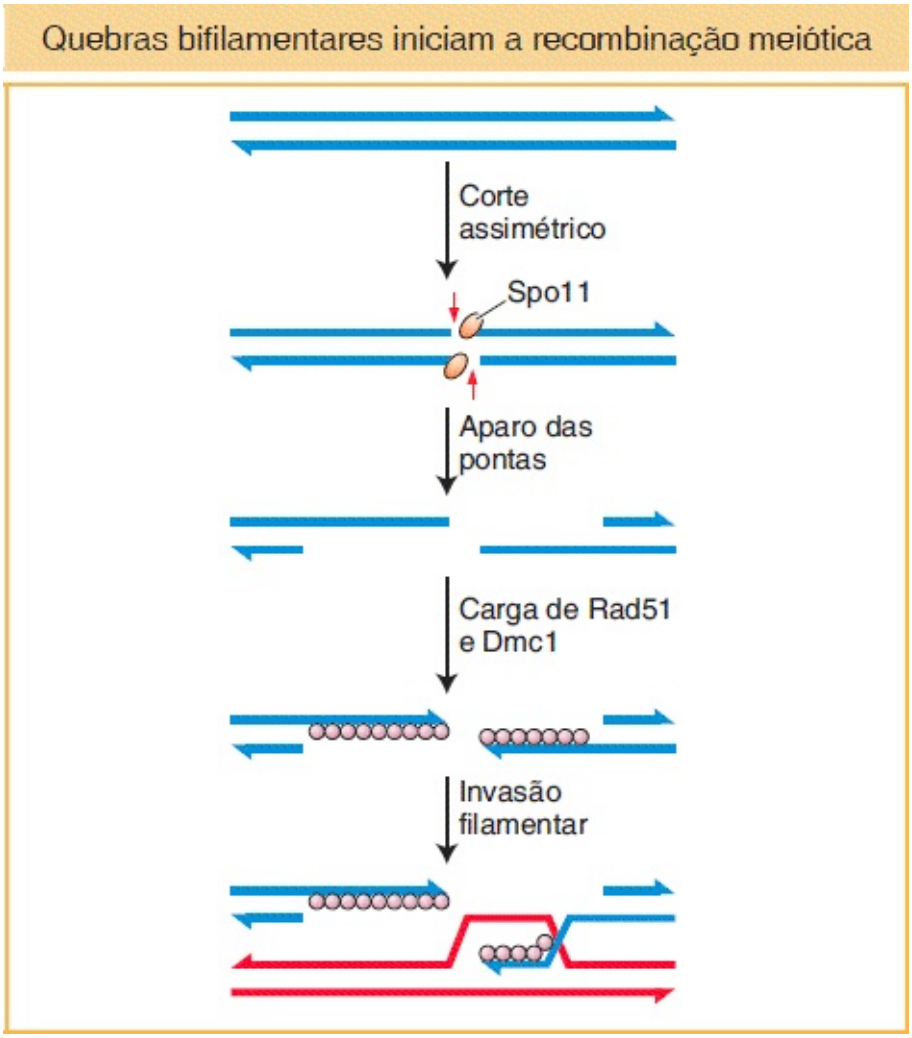


FIGURA 16.29 A recombinação meiótica é iniciada quando a enzima Spo11 realiza cortes desencontrados em um par de filamentos de DNA em uma cromátide.

CONCEITO-CHAVE A recombinação meiótica é iniciada pela enzima Spo11, que introduz cortes bifilamentares nos cromossomos após terem se

replicado, mas antes da separação dos homólogos.

16.5 Câncer | Uma importante consequência fenotípica da mutação

Por que tantos agentes mutagênicos causam câncer? Qual é a conexão entre o câncer e a mutação? Nesta seção, exploraremos a conexão mutação-câncer.

Tornou-se claro que praticamente todos os cânceres de células somáticas surgem em virtude de uma série de mutações especiais que se acumulam em uma célula. Algumas dessas mutações alteram a atividade de um gene; outras simplesmente eliminam a atividade gênica. As mutações que promovem o câncer estão classificadas em uma de algumas categorias importantes: aquelas que aumentam a capacidade de uma célula de proliferar; aquelas que diminuem a suscetibilidade de uma célula a uma via suicida, denominada **apoptose**; ou aquelas que aumentam a taxa de mutações geral da célula ou a sua longevidade, de modo que todas as mutações, incluindo aquelas que encorajam a proliferação ou a apoptose, apresentam maior probabilidade de ocorrência.

Como as células cancerosas diferem das células normais

Um tumor maligno, ou **câncer**, é um agregado de células, todas descendentes de uma célula aberrante inicial. Em outras palavras, todas as células malignas são membros de um único clone, até mesmo em cânceres avançados que apresentam múltiplos tumores em muitos locais no corpo. As células cancerosas diferem tipicamente de suas vizinhas normais em uma diversidade de características fenotípicas, tais como rápida taxa de divisão, capacidade de invadir novos territórios celulares, alta taxa metabólica e forma anormal. Por exemplo, quando as células de folhetos celulares epiteliais normais são semeadas em uma cultura celular, elas conseguem crescer apenas quando ancoradas à própria placa de cultura. Além disso, as células epiteliais normais em cultura se dividem apenas até formarem uma única camada contínua (Figura 16.30 A). Nesse ponto, de alguma maneira elas reconhecem que formaram uma única camada epitelial e

param de se dividir. Contrariamente, as células malignas derivadas de tecido epitelial continuam a proliferar, empilhando-se umas sobre as outras (Figura 16.30 B).

Claramente, os fatores que regulam a fisiologia celular normal foram alterados. Qual, então, é a causa subjacente do câncer? Muitos tipos celulares diferentes podem ser convertidos em um estado maligno. Existe um tema em comum? Ou cada um surge de modo diferente? Podemos pensar a respeito do câncer, de modo geral, como consequente ao acúmulo de múltiplas mutações em uma única célula, que causam a sua proliferação descontrolada. Algumas dessas mutações podem ser transmitidas a partir dos genitores por meio da linhagem germinativa. A maior parte surge *de novo* na linhagem somáticas de uma célula em particular.

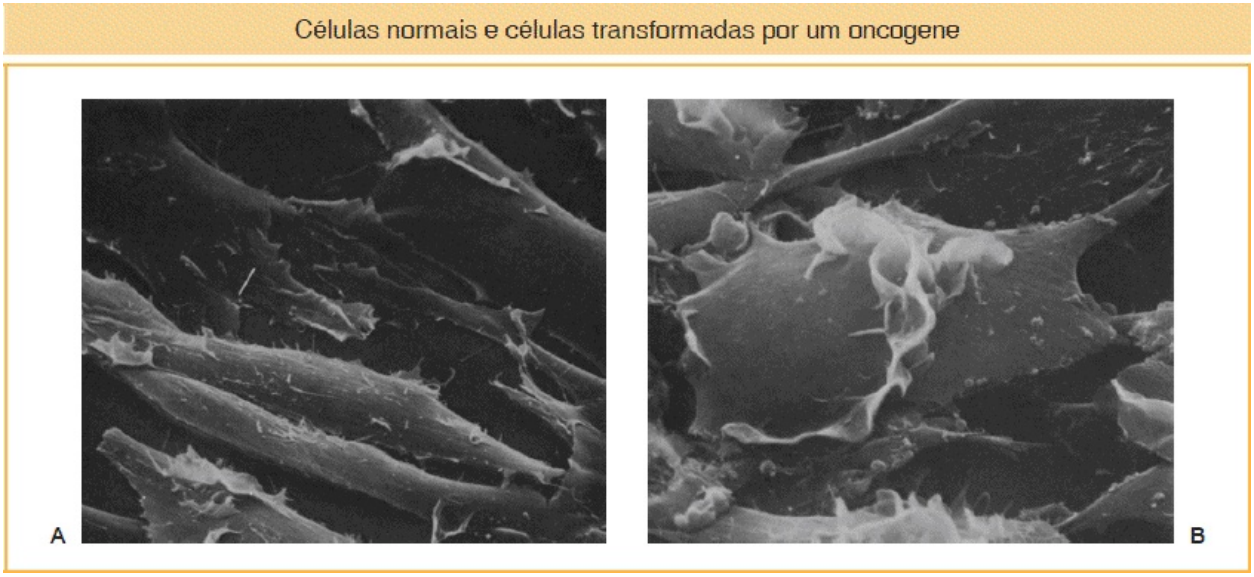


FIGURA 16.30 Micrografias de varredura eletrônica de (**A**) células normais e (**B**) células transformadas pelo vírus do sarcoma de Rous, que infecta as células com o oncogene *src*. **A.** Uma linhagem celular normal, denominada 3T3. Observe a estrutura de monocamada organizada das células. **B.** Uma derivativa transformada de 3T3. Observe como as células são mais arredondadas e estão empilhadas umas sobre as outras. (De Victor R. Ambros, Lan Bo Chen e John M. Buchanan, "Surface Ruffles as Markers for Studies of Cell Transformation by Rous Sarcoma Virus", *Proc. Nat. Acad. Sci. USA* 72, No. 8, 3144-3148, August 1975, *Cell Biology*, p. 3144, Figure 1A e 1B.)

Mutações em células cancerosas

Diversas linhas de evidências apontam para uma origem genética em relação à

transformação de células do estado benigno até o canceroso. Primeiramente, conforme já discutido neste capítulo, muitos agentes mutagênicos, tais como substâncias químicas e radiação, causam câncer, sugerindo que eles introduzam mutações nos genes. Em segundo lugar, e mais importante, foram identificadas mutações que frequentemente estão associadas a tipos de cânceres em particular.

Dois tipos gerais estão associados aos tumores: mutações em oncogenes e mutações em genes supressores de tumor. As mutações em **oncogenes** atuam na célula cancerosa como mutações dominantes de ganho de função (ver Capítulo 6 para uma discussão sobre as mutações dominantes). Essa declaração sugere duas características-chave das mutações em oncogenes. Primeiramente, as proteínas codificadas pelos oncogenes em geral são *ativadas* em células tumorais e, em segundo lugar, a mutação deve estar presente em apenas um alelo para contribuir para formação do tumor. O gene em sua forma normal e não mutado é denominado **proto-oncogene**.

As mutações nos **genes supressores de tumor** que promovem a formação de tumor são mutações recessivas de perda de função. Ou seja, esse tipo de mutação faz com que os produtos dos genes codificados percam uma grande parte de sua atividade ou toda ela (ou seja, a mutação é nula). Além disso, para que o câncer se desenvolva, a mutação deve estar presente em ambos os alelos do gene.

> CONCEITO-CHAVE Os oncogenes codificam formas mutadas de proteínas celulares normais, que resultam em mutações dominantes, normalmente em virtude de sua ativação inadequada. Contrariamente, os genes supressores de tumor codificam proteínas cuja perda de atividade pode contribuir para um estado canceroso. Como tal, são mutações recessivas.

Classes de oncogenes. Foi identificada aproximadamente uma centena de oncogenes diferentes. Como os seus correspondentes normais, os proto-oncogenes, atuam? Os proto-oncogenes em geral codificam uma classe de proteínas que estão ativas apenas quando os sinais regulatórios adequados possibilitam que eles sejam ativados. Muitos produtos de proto-oncogenes são

elementos em vias que induzem (controlam positivamente) o ciclo celular. Esses produtos incluem receptores de fatores de crescimento, proteínas de transdução de sinal e reguladores de transcrição. Outros produtos de proto-oncogenes atuam para inibir (controlar negativamente) a via apoptótica que destrói as células danificadas. Em ambos os tipos de mutação de oncogenes, a atividade da proteína mutante foi desacoplada de sua via reguladora normal, levando à sua expressão desregulada contínua. O produto proteico continuamente expresso de um oncogene é denominado **oncoproteína**. Diversas categorias de oncogenes foram identificadas de acordo com as diferentes vias nas quais as funções reguladoras foram desacopladas.

O oncogene *ras* pode ser utilizado para ilustrar o que ocorre quando um gene normal mantém uma mutação promotora de tumor. Conforme frequentemente é o caso, a alteração de uma proteína normal em uma oncoproteína envolve modificações estruturais da própria proteína – nesse caso, causada por uma simples mutação de ponto. Uma substituição de um único par de bases, que converte glicina em valina no aminoácido 12 da proteína Ras, por exemplo, cria a oncoproteína observada no câncer de bexiga humano (Figura 16.31 A). A proteína Ras normal é uma subunidade da proteína G que participa na transdução de sinal. Ela normalmente atua por meio da oscilação entre o estado ativo ligado ao GTP e o inativo ligado ao GDP. A mutação de sentido trocado no oncogene *ras* produz uma oncoproteína que sempre se liga ao GTP (Figura 16.31 B), até mesmo na ausência de sinais normais. Consequentemente, a oncoproteína Ras propaga continuamente um sinal que promove a proliferação celular.

Genes supressores de tumor. As funções normais dos genes supressores de tumor estão classificadas em categorias complementares àquelas dos proto-oncogenes (Tabela 16.1). Alguns genes supressores de tumor codificam reguladores negativos cuja função normal é inibir o ciclo celular. Outros codificam reguladores positivos que normalmente ativam a apoptose, ou a morte celular, de uma célula danificada. Outros ainda são participantes indiretos no câncer, com um papel normal no reparo de DNA danificado ou no controle da longevidade celular. Consideraremos um exemplo aqui.

Mutações no gene *p53* estão associadas a muitos tipos de tumores. De fato, estima-se que 50% dos tumores humanos apresentem ausência de um gene *p53* funcional. A proteína p53 ativa é um regulador transcricional que é ativado em resposta a dano no DNA. A p53 do tipo selvagem ativada atua em uma função dupla: previne a progressão do ciclo celular até que o dano ao DNA seja reparado e, sob algumas circunstâncias, induz apoptose. Se nenhum gene *p53* funcional está presente, o ciclo celular progride até mesmo se o DNA danificado não houver sido reparado. A progressão do ciclo celular para a mitose eleva a frequência geral de mutações, rearranjos cromossômicos e aneuploidia e, portanto, aumenta as chances de que surjam outras mutações que promovam a proliferação celular ou bloqueiem a apoptose.

Agora está claro que as mutações capazes de elevar a taxa de mutação são importantes contribuintes para a progressão de tumores em seres humanos. Essas mutações são recessivas nos genes supressores de tumor que normalmente atuam nas vias de reparo do DNA. Mutações nesses genes, portanto, interferem no reparo do DNA. Elas promovem o crescimento tumoral indiretamente, ao elevar a taxa de mutação, o que torna muito mais provável o surgimento de uma série de mutações em oncogenes e supressores de tumor, corrompendo a regulação normal do ciclo celular e a morte celular programada. Foram identificadas inúmeras das referidas mutações em genes supressores de tumor, incluindo algumas associadas a formas hereditárias de câncer em tecidos específicos. Exemplos são as mutações em *BRCA1* e *BRCA2* e o câncer de mama.

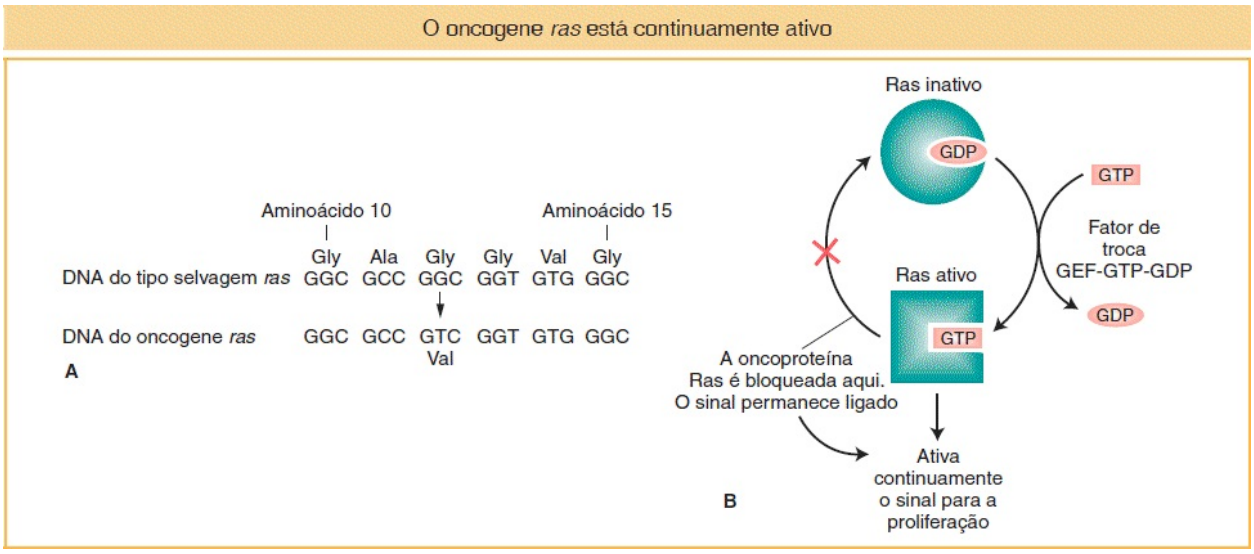


FIGURA 16.31 Formação e efeito da oncoproteína Ras. **A.** O oncogene *ras* difere do tipo selvagem em um único par de bases, produzindo uma oncoproteína Ras que difere do tipo selvagem em um aminoácido, na posição 12 na matriz de leitura aberta de *ras*. **B.** A oncoproteína Ras não consegue hidrolisar GTP em GDP. Em virtude deste defeito, a oncoproteína Ras permanece no complexo ativo Ras-GTP e ativa continuamente o sinal para a proliferação.

Tabela 16.1 Funções de proteínas do tipo selvagem e propriedades das mutações promotoras de tumor nos genes correspondentes.

Função da proteína do tipo selvagem	Propriedades das mutações promotoras de tumor
Promove a progressão do ciclo celular	Oncogene (ganho de função)
Inibe a progressão do ciclo celular	Mutação de supressor de tumor (perda de função)
Promove apoptose	Mutação de supressor de tumor (perda de função)
Inibe apoptose	Oncogene (ganho de função)

Promove o reparo de DNA	Mutação de supressor de tumor (perda de função)

CONCEITO-CHAVE Agentes mutagênicos podem causar alguns cânceres, tendo em vista que o câncer é, em parte, causado por versões mutantes de genes normais que levam ao crescimento descontrolado.

RESUMO

A alteração do DNA em um gene (mutação de ponto) em geral envolve um ou alguns poucos pares de bases. As substituições de pares de bases únicos podem criar códons de sentido trocado ou códons sem sentido (término de tradução). Uma purina substituída pela outra purina (ou uma pirimidina substituída pela outra pirimidina) é uma transição. Uma purina substituída por uma pirimidina (ou vice-versa) é uma transversão. Adições ou deleções (indels) de pares de bases únicos produzem mudança de matriz de leitura. Determinados genes humanos que contêm repetições de trinucleotídios — especialmente aqueles que são expressos no tecido neural — se tornam mutados por meio da expansão dessas repetições e, assim, podem causar doenças. A formação de repetições de monoaminoácidos nos polipeptídios codificados por esses genes com frequência é responsável pelos fenótipos mutantes.

As mutações podem ocorrer espontaneamente como um subproduto de processos celulares normais, tais como replicação do DNA ou metabolismo, ou podem ser induzidas por radiação mutagênica ou substâncias químicas. Os mutágenos com frequência resultam em um tipo específico de alteração em virtude de sua especificidade química. Por exemplo, alguns produzem exclusivamente transições $G \cdot C \rightarrow A \cdot T$; outros, exclusivamente mudanças de matriz de leitura.

Embora as mutações sejam necessárias para gerar diversidade, muitas mutações estão associadas a doenças genéticas hereditárias, tais como o xeroderma pigmentoso. Além disso, as mutações que ocorrem em células

somáticas são a fonte de muitos cânceres humanos. Foram desenvolvidas muitas vias biológicas para corrigir o amplo espectro de mutações espontâneas e induzidas. Algumas vias, tais como o reparo por excisão de base e nucleotídio e o reparo de malpareamento, utilizam as informações inerentes à complementaridade de bases para executar reparo livre de erro. Outras vias que utilizam polimerases de *bypass* para corrigir bases danificadas podem introduzir erros na sequência de DNA.

A correção de quebras bifilamentares é particularmente importante, tendo em vista que essas lesões podem levar a rearranjos cromossômicos desestabilizantes. A junção de extremidades não homólogas é uma via que une as extremidades quebradas, de modo que uma forquilha de replicação parada não resulte em morte celular. Nas células em replicação, as quebras bifilamentares podem ser reparadas de um modo livre de erro por meio da via de pareamento de filamento dependente de síntese, que utiliza a cromátide-irmã para reparar a quebra.

Centenas de quebras bifilamentares programadas iniciam o *crossing over* meiótico entre as cromátides não irmãs. Assim como outras quebras bifilamentares, as quebras meióticas devem ser processadas rapidamente e com eficiência para prevenir consequências sérias, tais como a morte celular e o câncer. O modo exato como esse reparo é realizado ainda está sendo explorado.

TERMOS-CHAVE

agente intercalar
análogo de base
apoptose
câncer
gene supressor de tumor
junção de extremidades não homólogas (NHEJ)
lesão espontânea
mudança de matriz de leitura
mutação de ponto
mutação de sentido trocado

mutação espontânea
mutação indel
mutação induzida
mutação sem sentido
mutação sinônima
mutagênese
mutágeno
oncogene
oncoproteína
pareamento de filamento dependente de síntese (SDSA)
plaqueamento em réplica
polimerase de *bypass* (translesão)
polimerase translesão (de *bypass*)
proto-oncogene
quebra bifilamentar
reparo de malpareamento
reparo por excisão de base
reparo por excisão de nucleotídio (NER)
reparo por excisão de nucleotídio genômico global (GG-NER)
reparo por excisão de nucleotídio acoplado à transcrição (TC-NER)
repetição de trinucleotídios
síndrome de Cockayne
síntese de DNA translesão
sistema SOS
sítio apurínico
substituição não conservativa
teste de Ames
teste de flutuação
transição
transversão
xeroderma pigmentoso (XP)

PROBLEMAS RESOLVIDOS

Problema resolvido 1. No Capítulo 9, aprendemos que os códons UAG e UAA são duas trincas sem sentido de término de cadeia. Com base na especificidade da aflatoxina B_1 e do etilmetanossulfonato (EMS), descreva se cada mutágeno seria capaz de reverter esses códons para o tipo selvagem.

Solução

O EMS induz primariamente transições G · C → A · T. Os códons UAG não podem ser revertidos para o tipo selvagem, tendo em vista que apenas a alteração de UAG → UAA seria estimulada pelo EMS e geraria um códon sem sentido (ocre). Os códons UAA não atuariam por meio do EMS. A aflatoxina B_1 induz primariamente transversões de G · C → T · A. Apenas a terceira posição dos códons UAG atuaria, resultando em uma alteração de UAG → UAU (no nível do mRNA), que produz tirosina. Portanto, se a tirosina for um aminoácido aceitável no sítio correspondente na proteína, a aflatoxina B_1 pode reverter os códons UAG. A aflatoxina B_1 não reverteria os códons UAA, tendo em vista que nenhum par de bases G · C aparece na posição correspondente no DNA.

Problema resolvido 2. Explique por que as mutações induzidas por acridinas no fago T4 ou por ICR-191 em bactérias não podem ser revertidas pela 5-bromouracila.

Solução

Acridinas e ICR-191 induzem mutações por meio da deleção ou da adição de um ou mais pares de bases, o que resulta em uma mudança da matriz de leitura. Entretanto, a 5-bromouracila induz mutações ao causar a substituição de uma base por outra. Essa substituição não consegue compensar a mudaça de matriz de leitura que resulta de ICR-191 e acridinas.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 16.3 A, qual é a consequência do novo sítio de corte 5′ na matriz de leitura aberta? Na 16.3 B, quão grande poderia ser o íntron para manter a matriz de leitura (digamos, entre 75 e 100 pb)?

2. Com a utilização da Figura 16.4 como um exemplo, compare a migração do RNA e da proteína em relação ao gene do tipo selvagem e a mutação demonstrada na Figura 16.3 B. Presuma que o íntron retido mantém a matriz de leitura.

3. No teste de Ames demonstrado na Figura 16.17, qual é o motivo para a adição do extrato hepático a cada amostra?

4. Com base no modo de ação da aflatoxina (Figura 16.16), proponha uma situação que explique a sua resposta no teste de Ames (Figura 16.18).

5. Na Figura 16.22, aponte a(s) proteína(s) mutante(s) em pacientes com síndrome de Cockayne. Qual(is) proteína(s) é/são mutante(s) em pacientes com XP? Como se acredita que essas diferentes mutações sejam responsáveis pelos diferentes sintomas da doença?

6. A proteína MutH corta o filamento recém-sintetizado (Figura 16.23). Como ela "sabe" qual é esse filamento?

7. Quais características da polimerase de *bypass* a tornam ideal para o seu papel na síntese translesão, demonstrada na Figura 16.24?

PROBLEMAS BÁSICOS

8. Considere as sequências do tipo selvagem e mutante a seguir:

Tipo selvagem	….CTTGCAAGCGAATC….
Mutante	….CTTGCTAGCGAATC….

A substituição demonstrada *parece* ter criado um códon de fim. Quais informações adicionais você necessita para estar confiante de que isso

tenha ocorrido?

9. Qual tipo de mutação está ilustrado pelas sequências a seguir (demonstradas como mRNA)?

Tipo selvagem	….5′ AAUCCUUACGGA 3′….
Mutante	….5′ AAUCCUACGGA 3′ ….

10. Uma mutação de sentido trocado de prolina para histidina pode ser realizada com mutágeno que causa uma transição G · C → A · T? O que dizer a respeito de uma mutação de sentido trocado de prolina para serina?

11. Por meio da substituição de par de bases, quais são todas as alterações sinônimas que podem ser produzidas com início no códon CGG?

12.

a. Quais são todas as transversões que podem ser realizadas com início no códon CGG?

b. Quais dessas transversões serão de sentido trocado? Você pode ter certeza?

13.

a. A acridina laranja é um mutágeno eficaz para a produção de alelos nulos por meio de mutação. Por que ela produz alelos nulos?

b. Um determinado composto semelhante à acridina dá origem apenas a inserções únicas. Uma mutação induzida com esse composto é tratada com o mesmo composto e são produzidos alguns revertentes. Como esse desfecho é possível?

14. Defenda a declaração: “O câncer é uma doença genética.”

15. Forneça um exemplo de um defeito de reparo do DNA que leve ao câncer.

16. No reparo de malpareamento em *E. coli*, apenas um erro no filamento recém-sintetizado é corrigido. Como a *E. coli* é capaz de reconhecer o

filamento recém-sintetizado? Por que essa capacidade faz sentido biológico?

17. Uma lesão mutacional resulta em uma sequência que contém um par de bases malpareado:

5′ AGCT**G**CCTT 3′
3′ ACG<u>A**T**G</u>GAA 5′
Códon

Se o reparo de malpareamento ocorre em qualquer direção, quais aminoácidos poderiam ser observados nesse sítio?

18. Sob quais circunstâncias poderia ser dito que a junção de extremidades não homólogas é propensa a erro?

19. Por que tantas substâncias químicas que apresentam resultado positivo no teste de Ames também são classificadas como carcinogênicas?

20. A proteína Spo11 está conservada em eucariotos. Você acredita que ela também esteja conservada em espécies bacterianas? Justifique a sua resposta.

21. Diferencie os elementos dos pares a seguir:
a. Transições e transversões.
b. Mutações sinônimas e neutras.
c. Mutações de sentido trocado e sem sentido.
d. Mudanças de matriz de leitura e mutações sem sentido.

22. Descreva duas lesões espontâneas que podem levar a mutações.

23. O que são polimerases de *bypass*? Como elas diferem das polimerases replicativas? Como as suas características especiais facilitam seu papel no reparo do DNA?

24. Em células adultas que pararam de se dividir, quais tipos de sistemas de reparo são possíveis?

25. Um determinado composto que é um análogo da base citosina pode se

tornar incorporado ao DNA. Ele normalmente forma pontes de hidrogênio com a citosina, mas com uma razoável frequência isomeriza-se em uma forma que faz pontes de hidrogênio com a timina. Você espera que esse composto seja mutagênico e, em caso afirmativo, quais tipos de alterações ele pode induzir no DNA?

26. Duas vias, a recombinação homóloga e a junção de extremidades não homólogas (NHEJ), podem reparar quebras bifilamentares no DNA. Se a recombinação homóloga é uma via livre de erros, enquanto NHEJ nem sempre é livre de erros, por que NHEJ é utilizada na maior parte das vezes em eucariotos?

27. Qual via de reparo reconhece o dano ao DNA durante a transcrição? O que ocorre se o dano não for reparado?

28. Onde em um gene uma inserção de 4 pb apresentaria o menor efeito sobre a expressão gênica?

a. Íntrons.
b. Éxons.
c. Regiões reguladoras.
d. Íntrons e éxons.

29. Qual das mutações gênicas a seguir apresenta maior probabilidade de provocar o impacto mais grave sobre a expressão gênica?

a. Uma mutação sem sentido no último éxon.
b. Uma mutação de ponto em um éxon.
c. Uma mutação de ponto no sítio doador de corte de um íntron.
d. Uma mutação de ponto na parte intermediária de um íntron.

30. Qual das seguintes não é possível?

a. Uma mutação não sinônima em um íntron.
b. Uma mutação não sinônima em um éxon.
c. Uma mutação indel em um íntron.
d. Uma mutação indel em um éxon.

31. Qual(is) da(s) seguintes está/estão associada(s) à mutação espontânea?

a. Uma ocorrência de câncer de pulmão em virtude de tabagismo.

b. Uma mutação sem sentido em um éxon, causada por um erro na replicação do DNA.

c. Uma mutação indel em um íntron, causada pelo deslize da replicação.

d. Uma mutação sem sentido em um éxon, causada por um erro na replicação do DNA e uma mutação indel em um íntron, causada pelo deslize da replicação.

32. Qual das declarações a seguir descreve melhor a via de reparo de malpareamento?

a. É parte da função de revisão 3′ para 5′ da DNA polimerase.

b. Atua após a replicação do DNA ao reconhecer pares de bases malpareadas.

c. É ativada por forquilhas de replicação paradas.

d. Está acoplada à transcrição.

PROBLEMAS DESAFIADORES

33.

a. Por que é impossível induzir mutações sem sentido (representadas no nível do mRNA pelas trincas UAG, UAA e UGA) por meio do tratamento de linhagens do tipo selvagem com mutágenos que causam apenas transições A · G → T · C no DNA?

b. A hidroxilamina (HA) causa apenas transições G · C → A · T no DNA. A HA produzirá mutações sem sentido em linhagens do tipo selvagem?

c. O tratamento com HA reverterá as mutações sem sentido?

34. Várias mutações de ponto auxotróficas em *Neurospora* são tratadas com diversos agentes para verificar se ocorrerá a reversão. Foram obtidos os resultados a seguir (um sinal de mais indica reversão; HA causa apenas transições G · C → A · T).

Mutante	5-BU	HA	Proflavina	Reversão espontânea

1	—	—	—	—
2	—	—	+	+
3	+	—	—	+
4	—	—	—	+
5	+	+	—	+

a. Em relação a cada um dos cinco mutantes, descreva a natureza do evento de mutação original (não a reversão) no nível molecular. Seja tão específico quanto possível.

b. Em relação a cada um dos cinco mutantes, denomine um possível mutágeno que possa ter causado o evento de mutação original. (Mutação espontânea não é uma resposta aceitável.)

c. No experimento de reversão em relação ao mutante 5, é obtido um derivativo prototrófico particularmente interessante. Quando esse tipo é cruzado com uma linhagem do tipo selvagem padrão, a progênie é composta por 90% de prototróficos e 10% auxotróficos. Forneça uma explicação completa para esses resultados, incluindo um motivo preciso para as frequências observadas.

35. Você está utilizando a nitrosoguanidina para "reverter" alelos mutantes *nic-2* (que necessitam de nicotinamida) em *Neurospora*.

Você trata as células, plaqueia-as em um meio sem nicotinamida e procura por colônias prototróficas. Você obtém os resultados a seguir em relação a dois alelos mutantes. Explique esses resultados no nível molecular e indique como você testaria as suas hipóteses.

a. Com o alelo 1 de *nic-2*, você não obtém prototróficos.

b. Com o alelo 2 de *nic-2*, você obtém três colônias prototróficas A, B e C, e cruza cada uma separadamente com uma linhagem do tipo selvagem. A partir do cruzamento prototrófico A × tipo selvagem, você obtém uma progênie de 100, todos prototróficos. A partir do cruzamento prototrófico B × tipo selvagem, você obtém uma progênie de 100, da qual 78 são

prototróficos e 22 necessitam de nicotinamida. A partir do cruzamento prototrófico C × tipo selvagem, você obtém uma progênie de 1.000, da qual 996 são prototróficos e 4 necessitam de nicotinamida.

36. Você está trabalhando com um mutágeno recentemente descoberto e deseja determinar a alteração de base que ele introduz no DNA. Até agora, você determinou que o mutágeno altera quimicamente uma única base de tal modo que as suas propriedades de pareamento de base são alteradas permanentemente. Para determinar a especificidade da alteração, você examina as alterações dos aminoácidos que ocorrem após a mutagênese. Um exemplo do que você encontra está demonstrado aqui:

Original: Gln-His-Ile-Glu-Lis

Mutante: Gln-His-Met-Glu-Lis

Original: Ala-Val-Asn-Arg

Mutante: Ala-Val-Ser-Arg

Original: Arg-Ser-Leu

Mutante: Arg-Ser-Leu-Trp-Lis-Tre-Fen

Qual é a especificidade da alteração de base do mutágeno?

37. Agora você encontra um mutante adicional a partir do experimento no Problema 31:

Original: Ile-Leu-His-Gln

Mutante: Ile-Pro-His-Gln

A especificidade da alteração de base na sua resposta ao Problema 31 poderia ser responsável por essa mutação? Por que sim ou por que não?

38. Você é um perito em mecanismos de reparo do DNA. Você recebe uma amostra de uma linhagem celular humana derivada de uma mulher que apresenta sintomas de xeroderma pigmentoso. Você determina que ela

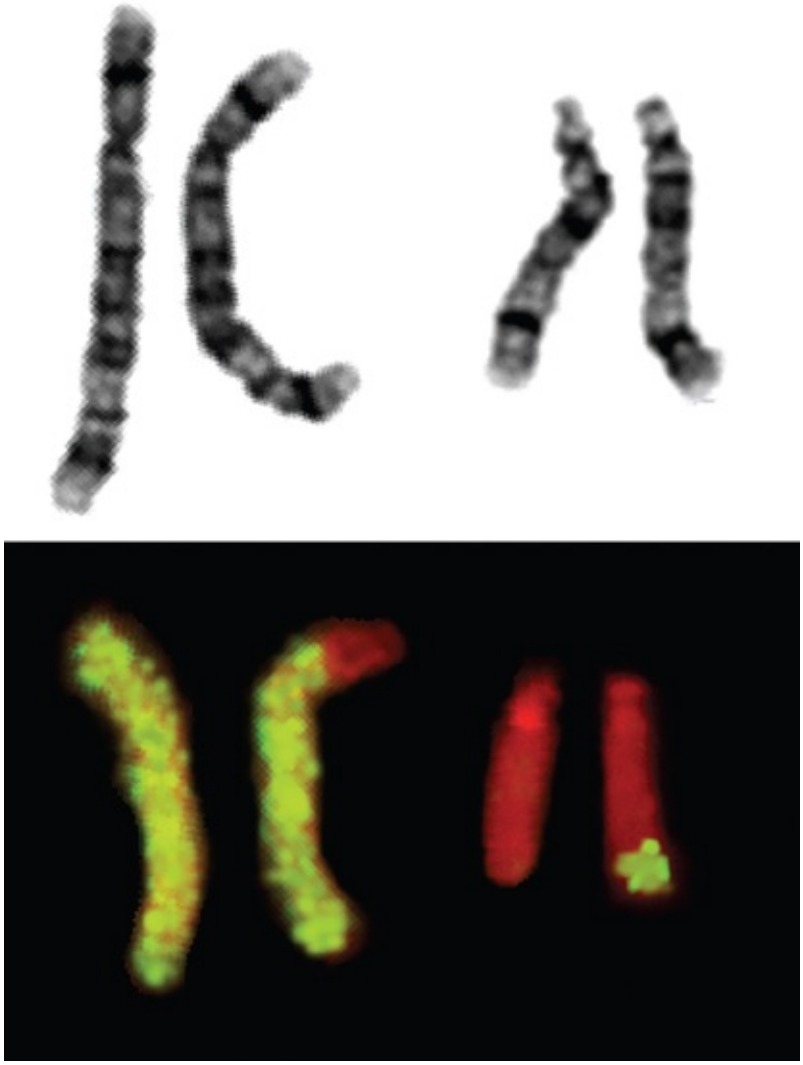
apresenta uma mutação em um gene que não foi associada anteriormente ao XP. Como isso é possível?

39. O ozônio (O_3) é um componente de ocorrência natural importante na nossa atmosfera, na qual ele forma uma camada que absorve a radiação UV. Foi descoberto um buraco na camada de ozônio na década de 1970 sobre a Antártica e a Austrália. O buraco aparece sazonalmente e observou-se que ocorre em virtude da atividade humana. Especificamente, o ozônio é destruído por uma classe de substâncias químicas (denominadas CFC, em referência a clorofluorocarbonos), que são observadas em refrigerantes, sistemas de ar-condicionado e aerossóis.

Como um cientista que trabalha com mecanismos de reparo do DNA, você descobre que tem ocorrido um aumento significativo no câncer de pele nas comunidades praianas na Austrália. Um amigo repórter jornalístico oferece deixar que você publique um breve comentário (um parágrafo), no qual você deve descrever a possível conexão entre o buraco de ozônio e o aumento dos cânceres de pele. Com base no que você aprendeu a respeito do reparo do DNA neste capítulo, escreva um parágrafo que explique essa conexão.

CAPÍTULO 17

Alterações Cromossômicas em Grande Escala



Uma translocação recíproca demonstrada por coloração de cromossomos. Uma suspensão de cromossomos de muitas células passa por um dispositivo eletrônico que classifica os cromossomos por tamanho. O DNA é extraído de cromossomos individuais, desnaturado, ligado a um de diversos corantes fluorescentes, e em seguida adicionado a cromossomos parcialmente desnaturados em uma lâmina. O DNA fluorescente

“encontra” o seu próprio cromossomo e se liga ao longo de seu comprimento por meio da complementaridade de bases, “pintando-o”. Neste exemplo, foram utilizados um corante vermelho e um verde para pintar cromossomos diferentes. A figura demonstra preparações não coloridas (*acima*) e coloridas (*abaixo*). A preparação colorida demonstra um cromossomo verde normal, um vermelho normal e dois que apresentam segmentos trocados. (Addenbrookes Hospital/Science Source.)

TÓPICOS

17.1 Alterações no número de cromossomos
17.2 Alterações na estrutura dos cromossomos
17.3 Incidência geral de mutações cromossômicas humanas

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Distinguir entre os principais tipos de mutações cromossômicas no nível citológico
- Deduzir as configurações de pareamento meiótico em relação a todas as principais mutações cromossômicas
- Prever as razões da progênie de heterozigotos autopoliploides específicos em relação ao um ou mais genes*
- Desenhar cruzamentos para sintetizar um alotetraploide
- Prever o desfecho da não disjunção meiótica na primeira e na segunda divisões*
- Identificar um aneuploide com a utilização de critérios genéticos*
- Prever as razões na progênie de aneuploides específicos*
- Distinguir entre os principais tipos aneuploides humanos
- Na análise da progênie, diagnosticar a presença de um dos principais tipos de rearranjos cromossômicos (translocações, inversões, deleções, duplicações)*
- Em um cruzamento envolvendo um rearranjo específico conhecido, prever a herança de genes ligados e não ligados ao rearranjo*
- Prever os padrões de expressão de genes possivelmente afetados pela

variegação por efeito de posição.

**No caso de mutações cromossômicas, a análise da progênies envolve a análise de padrões de um ou mais dos seguintes: esterilidade, letalidade e proporções fenotípicas dos genes heterozigotos nos cruzamentos.*

Um casal jovem está planejando ter filhos. O marido sabe que sua avó teve um filho com síndrome de Down em um segundo casamento. A síndrome de Down é um conjunto de distúrbios físicos e mentais causado pela presença de um cromossomo 21 extra (Figura 17.1). Não se dispõe do nascimento, que ocorreu no início do século 20, mas o casal não conhece outros casos de síndrome de Down em suas famílias.

O casal ouviu dizer que a síndrome de Down resulta de um raro erro ao acaso na produção do zigoto e, portanto, acredita que apresenta apenas uma chance baixa de ter uma criança com a doença. Eles decidem ter filhos. Seu primeiro filho não é afetado, mas a próxima concepção resulta em aborto espontâneo e seu segundo filho nasce com síndrome de Down. O fato de terem um filho com síndrome de Down é uma coincidência ou uma conexão entre a constituição genética do pai da criança e a de sua avó levou ambos a terem filhos com síndrome de Down? O aborto espontâneo tem alguma relação? Quais testes podem ser necessários para investigar essa situação? A análise de tais questões é o tópico deste capítulo.

Verificamos em todo este livro que as mutações gênicas são uma fonte importante de alterações na sequência genômica. Entretanto, o genoma também pode ser remodelado em maior escala por meio de alterações na estrutura cromossômica, ou por meio de alterações no número de cópias dos cromossomos em uma célula. Essas variações em grande escala são denominadas **mutações cromossômicas**, para que sejam distinguidas das mutações gênicas. Em termos mais amplos, as mutações gênicas são definidas como alterações que ocorrem dentro de um gene, enquanto as mutações cromossômicas são alterações em uma região do cromossomo que engloba múltiplos genes. As mutações gênicas nunca são detectáveis microscopicamente; um cromossomo que contém uma mutação

gênica, ao microscópio, aparenta ser igual a um que carreia o alelo do tipo selvagem. Contrariamente, muitas mutações cromossômicas podem ser detectadas por microscopia, por análise genética ou molecular, ou por uma combinação de todas as técnicas. As mutações cromossômicas foram mais bem-caracterizadas em eucariotos e todos os exemplos neste capítulo são desse grupo.

Criança com síndrome de Down

FIGURA 17.1 A síndrome de Down resulta de uma cópia extra do cromossomo 21. (Terry Harris/Rex Features/Associated Press.)

As mutações cromossômicas são importantes a partir de diversas perspectivas biológicas. Primeiramente, elas podem ser fontes de percepção sobre o modo como os genes atuam em conjunto em uma escala genômica. Em segundo lugar, elas revelam diversas características importantes da meiose e da arquitetura

cromossômica. Em terceiro lugar, elas constituem ferramentas úteis para a manipulação genômica experimental. Em quarto lugar, elas são fontes de percepção a respeito dos processos evolutivos. Em quinto lugar, as mutações cromossômicas são observadas regularmente em seres humanos e algumas dessas mutações causam doenças genéticas.

Muitas mutações cromossômicas causam anormalidades na célula e na função do organismo. A maior parte dessas anormalidades tem origem em alterações no *número* ou na *posição dos genes*. Em alguns casos, uma mutação cromossômica resulta da quebra do cromossomo. Se a quebra ocorre dentro de um gene, o resultado é a *ruptura* funcional daquele gene.

Para os nossos objetivos, dividiremos as mutações cromossômicas em dois grupos: alterações no *número* de cromossomos e alterações na *estrutura* cromossômica. Esses dois grupos representam dois tipos de eventos fundamentalmente diferentes. As alterações no número de cromossomos não estão associadas a alterações estruturais de quaisquer das moléculas de DNA da célula. Em vez disso, é o *número* dessas moléculas de DNA que está alterado e essa alteração no número é a base dos seus efeitos genéticos. As alterações na estrutura cromossômica, por outro lado, resultam em novos arranjos de sequência em uma ou mais duplas-hélices de DNA. Esses dois tipos de mutações cromossômicas estão ilustrados na Figura 17.2, que é um resumo dos tópicos deste capítulo. Iniciamos explorando a natureza e as consequências das alterações no número de cromossomos.

17.1 Alterações no número de cromossomos

Na genética como um todo, poucos tópicos afetam as questões humanas tão diretamente quanto as alterações no número de cromossomos presentes em nossas células. Em primeiro lugar está o fato de que um grupo de distúrbios genéticos comuns resulta da presença de um número anormal de cromossomos. Embora esse grupo de distúrbios seja pequeno, ele é responsável por uma grande proporção dos problemas de saúde geneticamente determinados que afetam os seres

humanos. Também de relevância para os seres humanos é o papel das mutações cromossômicas na agricultura: cultivadores de plantas manipularam rotineiramente o número de cromossomos para melhorar os cultivos agrícolas comercialmente importantes.

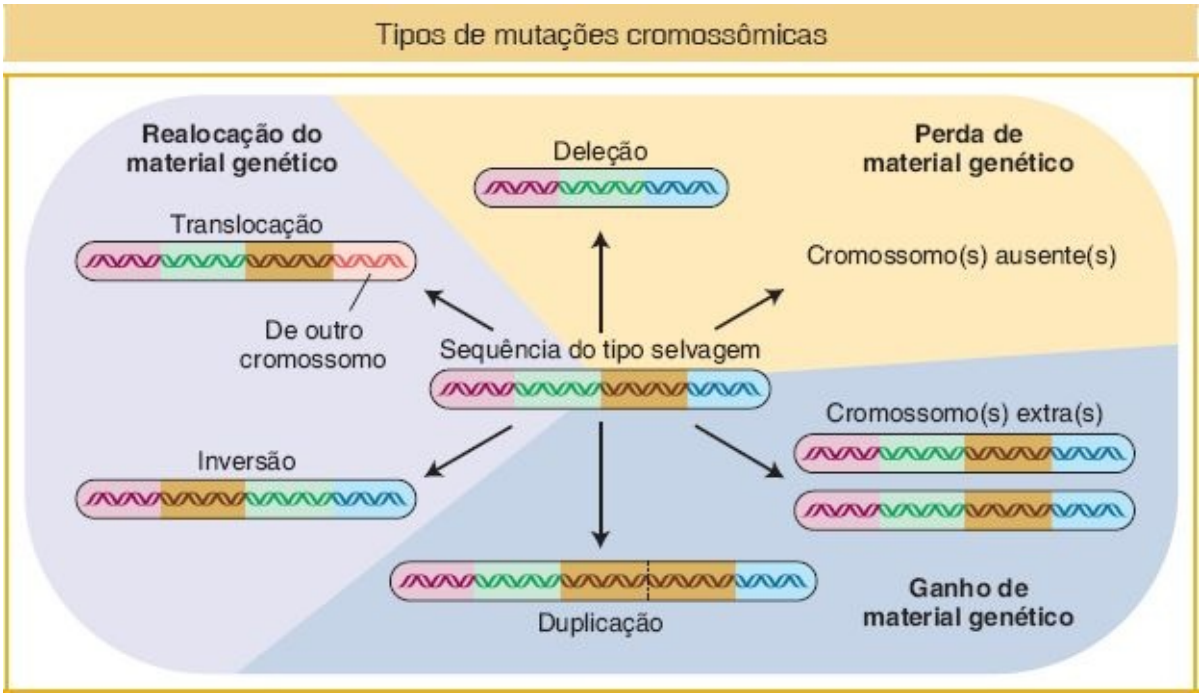


FIGURA 17.2 A ilustração está dividida em três regiões coloridas para ilustrar os principais tipos de mutações cromossômicas que podem ocorrer: a perda, o ganho ou a realocação de cromossomos inteiros ou de segmentos cromossômicos. O cromossomo do tipo selvagem está demonstrado ao centro.

As alterações no número de cromossomos são de dois tipos básicos: alterações em conjuntos *completos* de cromossomos, que resultam em uma condição denominada *euploidia aberrante* e alterações em *partes* dos conjuntos de cromossomos, que resultam em uma condição denominada *aneuploidia*.

Euploidia aberrante

Os organismos com múltiplos do conjunto cromossômico básico (genoma) são denominados **euploides**. Você aprendeu, nos capítulos anteriores, que eucariotos conhecidos, tais como plantas, animais e fungos, carreiam em suas células um conjunto de cromossomos (haploidia), ou dois conjuntos de cromossomos

(diploidia). Nessas espécies, tanto o estado haploide quanto o diploide são casos de euploidia normal. Os organismos que apresentam mais ou menos do que o número normal do conjunto cromossômico são euploides aberrantes. Os **poliploides** são organismos que apresentam mais de dois conjuntos de cromossomos. Eles podem ser representados por 3*n* (**triploide**), 4*n* (**tetraploide**), 5*n* (**pentaploide**), 6*n* (**hexaploide**) e assim por diante. (O número de conjuntos de cromossomos é denominado ploidia ou nível de ploidia.) Um membro de uma espécie normalmente diploide que apresenta apenas um conjunto de cromossomos (*n*) é denominado **monoploide**, para que seja distinguido de um membro de uma espécie normalmente haploide (também *n*). Exemplos dessas condições estão demonstrados nas quatro primeiras linhas da Tabela 17.1.

Monoploides. Machos de abelhas, vespas e formigas são monoploides. Nos ciclos de vida normais desses insetos, os machos se desenvolvem por meio de **partenogênese** (o desenvolvimento de um tipo especializado de ovo não fertilizado em um embrião sem a necessidade de fertilização). Entretanto, na maior parte das outras espécies, zigotos monoploides falham em se desenvolver. O motivo é que praticamente todos os membros de uma espécie diploide carreiam um número de mutações recessivas deletérias, denominadas em conjunto **carga genética**. Os alelos recessivos deletérios são mascarados por alelos do tipo selvagem na condição diploide, mas são automaticamente expressos em um monoploide derivado de um diploide. Os monoploides que chegam a se desenvolver até os estágios avançados são anormais. Se eles sobrevivem até a fase adulta, suas células germinativas não conseguem entrar em meiose normalmente, tendo em vista que os cromossomos não apresentam parceiros de pareamento. Portanto, os monoploides são caracteristicamente estéreis. (Nos machos de abelhas, vespas e formigas, não há meiose; nesses grupos, os gametas são produzidos por *mitose*.)

Tabela 17.1 Constituições cromossômicas em um organismo normalmente diploide com três cromossomos (identificados como A, B e C) no conjunto básico.

Nome	Designação	Constituição	Número de cromossomos
Euploides			
Monoploide	*n*	A B C	3
Diploide	2*n*	AA BB CC	6
Triploide	3*n*	AAA BBB CCC	9
Tetraploide	4*n*	AAAA BBBB CCCC	12
Aneuploides			
Monossômico	2*n* — 1	A BB CC	5
		AA B CC	5
		AA BB C	5
Trissômico	2*n* + 1	AAA BB CC	7
		AA BBB CC	7
		AA BB CCC	7

Poliploides. A poliploidia é muito comum em plantas, porém é mais rara em animais (por motivos que consideraremos posteriormente). De fato, um aumento no número de conjuntos cromossômicos tem sido um fator importante na origem de novas espécies de plantas. A evidência desse benefício é que, acima de um número haploide de aproximadamente 12, números pares de cromossomos são muito mais comuns do que números ímpares. Esse padrão é uma consequência da origem poliploide de muitas espécies de plantas, tendo em vista que a duplicação e a reduplicação de um número podem dar origem apenas a números pares. Espécies de animais não demonstram tal distribuição, em virtude da relativa raridade de animais poliploides.

Em euploides aberrantes, com frequência existe uma correlação entre o número de cópias do conjunto cromossômico e o tamanho do organismo. Um organismo tetraploide, por exemplo, tipicamente aparenta ser muito semelhante ao seu correspondente diploide em suas proporções, a não ser pelo fato de que o tetraploide é maior, no todo e em suas partes componentes. Quanto mais alto o nível de ploidia, maior o tamanho do organismo (Figura 17.3).

> CONCEITO-CHAVE Poliploides com frequência são maiores e apresentam partes componentes maiores do que os seus correlatos diploides.

No âmbito dos poliploides, devemos distinguir entre os **autopoliploides**, que apresentam múltiplos conjuntos cromossômicos que se originam dentro de uma espécie, e os **alopoliploides**, que apresentam conjuntos de duas ou mais espécies diferentes. Os alopoliploides são formados apenas entre espécies relacionadas de modo próximo; entretanto, os diferentes conjuntos cromossômicos são apenas **homeólogos** (parcialmente homólogos), não totalmente homólogos como são em autopoliploides.

Autopoliploides. Os triploides (3*n*) normalmente são autopoliploides. Eles surgem espontaneamente na natureza, mas podem ser obtidos por geneticistas a

partir do cruzamento de um 4*n* (tetraploide) com um 2*n* (diploide). Os gametas 2*n* e *n* produzidos pelo tetraploide e pelo diploide, respectivamente, unem-se para formar um triploide 3*n*. Os triploides são caracteristicamente estéreis. O problema (que também é verdadeiro em relação aos monoploides) está na presença de cromossomos não pareados na meiose. Os mecanismos moleculares para a sinapse, ou pareamento verdadeiro, ditam que, em um triploide, o pareamento pode ocorrer apenas entre dois dos três cromossomos de cada tipo (Figura 17.4). Os homólogos pareados (**bivalentes**) segregam-se para polos opostos, mas os homólogos não pareados (**univalentes**) passam para qualquer polo aleatoriamente. Em um **trivalente**, um grupo pareado de três, os centrômeros pareados segregam-se como um bivalente e o não pareado, como um univalente. Essas segregações ocorrem em relação a cada trio cromossômico; assim, em relação a qualquer tipo cromossômico, o gameta pode receber um ou dois cromossomos. É improvável que um gameta receba *dois* de *cada* tipo cromossômico, ou que receba *um* de *cada* tipo cromossômico. Portanto, a probabilidade é de que os gametas apresentem números de cromossomos intermediários entre o número haploide e o diploide; tais genomas são de um tipo denominado **aneuploide** (“não euploide”).

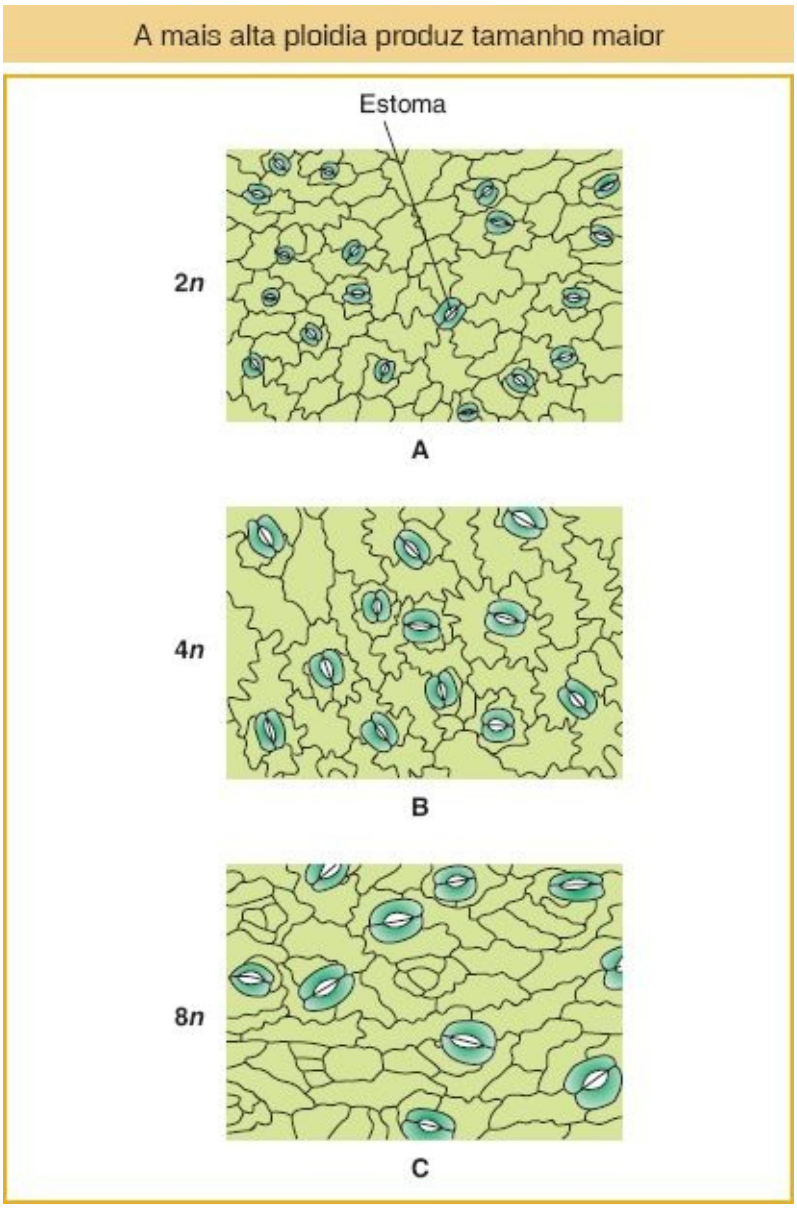


FIGURA 17.3 Células epidérmicas de folhas de plantas do tabaco com ploidia crescente. O tamanho da célula aumenta com o aumento da ploidia, o que é particularmente evidente no tamanho do estoma. **A.** Diploide. **B.** Tetraploide. **C.** Octoploide.

Os gametas aneuploides em geral não dão origem a descendência viável. Em plantas, os grãos de pólen aneuploides em geral são inviáveis e, portanto, incapazes de fertilizar o gameta feminino. Em qualquer organismo, os zigotos que podem ter origem a partir da fusão de um gameta haploide e um aneuploide serão, eles próprios, aneuploides e tipicamente esses zigotos também são inviáveis. Examinaremos o motivo subjacente à inviabilidade dos aneuploides quando

considerarmos o balanço gênico posteriormente no capítulo.

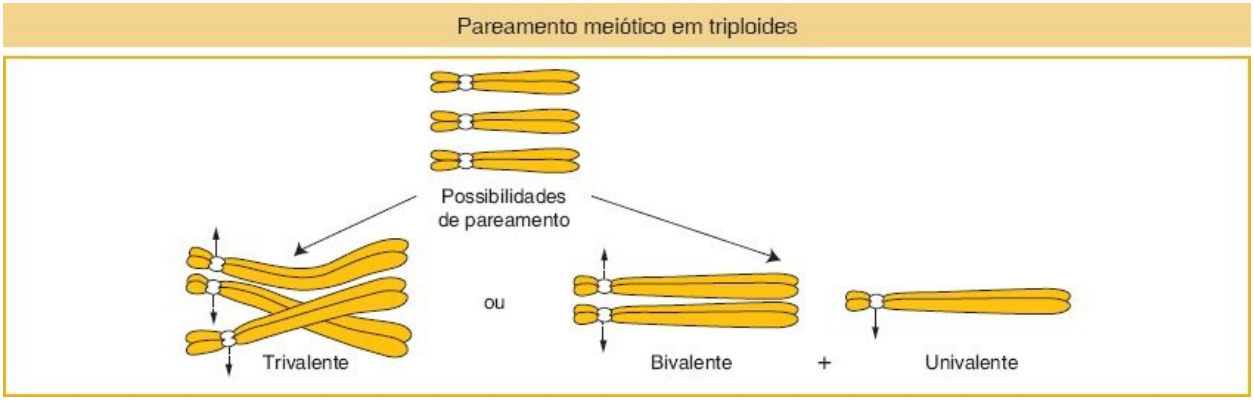


FIGURA 17.4 Os três cromossomos homólogos de um triploide podem parear de dois modos na meiose, como um trivalente ou como um bivalente mais um univalente.

> CONCEITO-CHAVE Os poliploides com números ímpares de conjuntos cromossômicos, tais como os triploides, são estéreis ou altamente inférteis, porque seus gametas e sua descendência são aneuploides.

Os autotetraploides têm origem a partir da duplicação de um complemento *2n* para 4*n*. Essa duplicação pode ocorrer espontaneamente, mas também pode ser induzida artificialmente por meio da aplicação de agentes químicos que interrompem a polimerização dos microtúbulos. Conforme declarado no Capítulo 2, a segregação cromossômica é movida pelas fibras do fuso, que são polímeros da proteína tubulina. Portanto, a interrupção da polimerização dos microtúbulos bloqueia a segregação dos cromossomos. O tratamento químico normalmente é aplicado ao tecido somático durante a formação das fibras do fuso em células que estão sofrendo divisão. O tecido poliploide resultante (tal como um ramo poliploide de uma planta) pode ser detectado por meio do exame dos cromossomos corados do tecido ao microscópio. Tal ramo pode ser removido e utilizado como uma muda para gerar uma planta poliploide, ou deixar que produza flores, as quais, quando autofecundadas, produzirão descendência poliploide. Um agente antitubulina comumente utilizado é a colchicina, um alcaloide extraído do

açafrão-do-prado. Em células tratadas com colchicina, a fase S do ciclo celular ocorre, mas não há segregação dos cromossomos ou divisão celular. Na medida em que a célula tratada entra na telófase, ocorre a formação de uma membrana nuclear ao redor do conjunto inteiro de cromossomos duplicados. Portanto, o tratamento de células diploides (2*n*) com a colchicina durante um ciclo celular leva à formação de tetraploides (4*n*) com exatamente quatro cópias de cada tipo de cromossomo (Figura 17.5). O tratamento durante um ciclo celular adicional produz octoploides (8*n*) e assim por diante. Esse método funciona com células de plantas e de animais, mas em geral as plantas aparentam ser muito mais tolerantes à poliploidia. Observe que todos os alelos no genótipo são duplicados. Portanto, se uma célula diploide de genótipo *A/a; B/b* for duplicada, o autotetraploide resultante será de genótipo *A/A/a/a*; *B/B/b/b*.

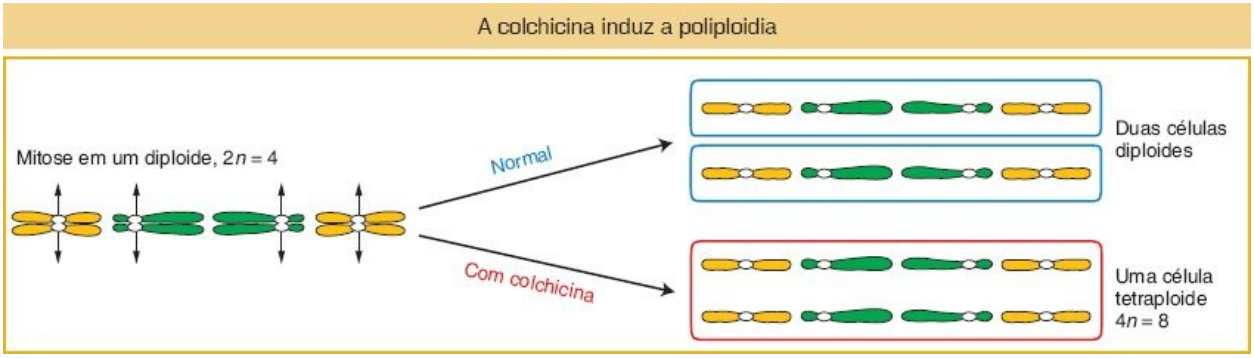


FIGURA 17.5 A colchicina pode ser aplicada para gerar um tetraploide a partir de um diploide. A colchicina adicionada às células mitóticas durante a metáfase e a anáfase interrompe a formação das fibras do fuso, impedindo a migração das cromátides após a divisão do centrômero. É criada uma única célula, que contém pares de cromossomos idênticos que são homozigotos em todos os *loci*.

Em virtude de quatro ser um número par, os autotetraploides podem apresentar uma meiose regular, embora esse resultado nem sempre ocorra. O fator crucial é o modo como os quatro cromossomos de cada par segregam-se. Existem diversas possibilidades, conforme demonstrado na Figura 17.6. Se os cromossomos parearem como bivalentes ou quadrivalentes, segregam-se normalmente, produzindo gametas diploides. A fusão dos gametas na fertilização regenera o estado tetraploide. Se trivalentes são formados, a segregação leva a gametas aneuploides não funcionais e, portanto, à esterilidade.

Quais razões genéticas são produzidas por um autotetraploide? Presuma, para simplificar, que o tetraploide forma apenas bivalentes. Se iniciarmos com uma planta tetraploide *A/A/a/a* e a autofecundarmos, qual proporção da progênie será *a/a/a/a*? Primeiramente, precisamos deduzir a frequência de gametas *a/a* em virtude de esse tipo ser o único que pode produzir um homozigoto recessivo. Os gametas *a/a* apenas podem surgir se ambos os pareamentos forem de *A* com *a* e, então, ambos os alelos *a* devem segregar para o mesmo polo. Utilizaremos o experimento hipotético a seguir para calcular as frequências dos possíveis desfechos. Considere as opções a partir do ponto de vista de um dos cromossomos *a* diante das opções de pareamento com o outro cromossomo *a*, ou com um dos dois cromossomos *A*; se o pareamento for aleatório, existe uma chance de dois terços de que ele irá parear com um cromossomo *A*. Se ele parear, então o pareamento dos dois cromossomos remanescentes necessariamente será de *A* com *a*, tendo em vista que aqueles são os únicos cromossomos remanescentes. Com esses dois pareamentos de *A* com *a*, existem duas segregações igualmente prováveis e, em geral, um quarto dos produtos conterá ambos os alelos *a* em um polo. Portanto, a probabilidade de um gameta *a/a* será 2/3 × 1/4 = 1/6. Portanto, se os gametas parearem aleatoriamente, a probabilidade de um zigoto *a/a/a/a* será de 1/6 × 1/6 = 1/36 e, por subtração, a probabilidade de *A/—/—/—* será de 35/36. Portanto, é esperada uma razão fenotípica de 35:1.

> CONCEITO-CHAVE Se poliploides sofrerem ordenadamente pareamento meiótico de dois centrômeros, podem resultar razões fenotípicas não mendelianas padrão.

Alopoliploides. Um alopoliploide é uma planta híbrida de duas ou mais espécies, contendo duas ou mais cópias de cada genoma incluído. O alopoliploide modelo foi um alotetraploide sintetizado por Georgi Karpechenko em 1928. Ele desejava produzir um híbrido fértil que apresentasse as folhas do repolho (*Brassica*) e as raízes do rabanete (*Raphanus*), tendo em vista que elas eram as partes importantes de cada planta em termos agrícolas. Cada uma dessas duas espécies

apresenta 18 cromossomos e, assim, $2n_1 = 2n_2 = 18$ e $n_1 = n_2 = 9$. As espécies são relacionadas de modo próximo o suficiente para possibilitar o cruzamento. A fusão de um gameta $n1$ e um $n2$ produziu um indivíduo de progênie híbrida viável, de constituição $n_1 + n_2 = 18$. Entretanto, esse híbrido era funcionalmente estéril, tendo em vista que os 9 cromossomos do genitor repolho eram suficientemente diferentes dos cromossomos do rabanete, de modo que os pares não realizavam sinapse, nem se segregavam normalmente na meiose e, portanto, o híbrido não poderia produzir gametas funcionais.

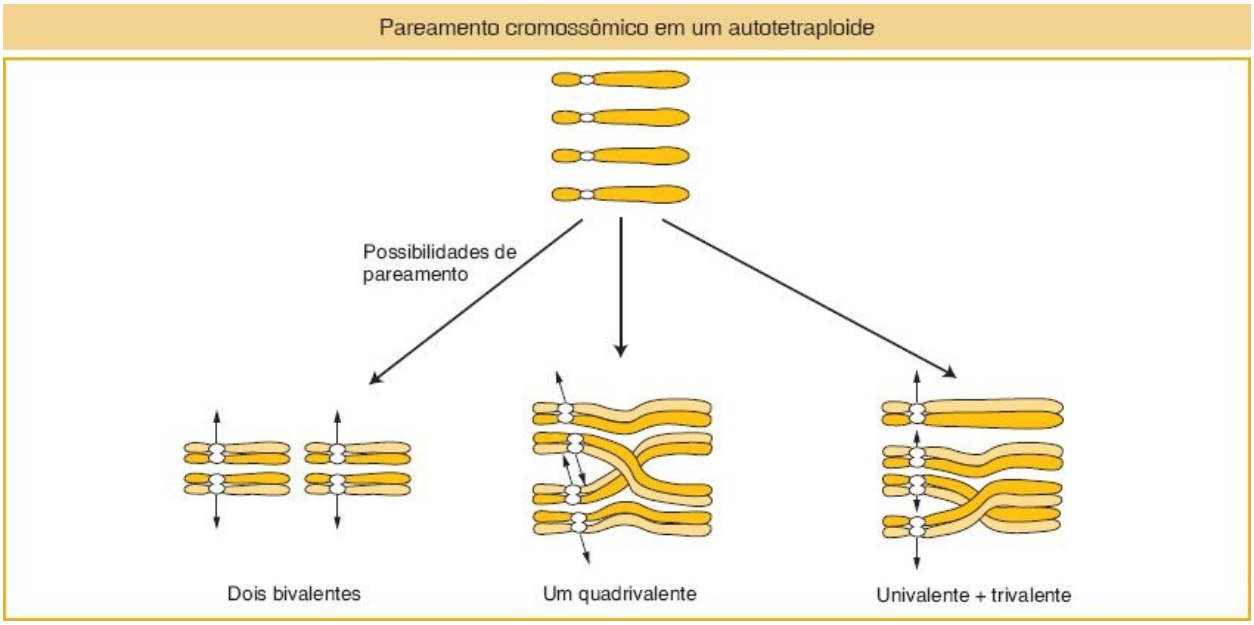


FIGURA 17.6 Existem três possibilidades de pareamento diferentes na meiose em tetraploides. Os quatro cromossomos homólogos podem parear-se como dois bivalentes ou como um quadrivalente, e cada um pode produzir gametas funcionais. Uma terceira possibilidade, um trivalente mais um univalente, produz gametas não funcionais.

Finalmente, uma parte da planta híbrida produziu algumas sementes. Após o plantio, essas sementes produziram indivíduos férteis com 36 cromossomos. Todos esses indivíduos eram alopoliploides. Aparentemente eles derivaram da duplicação cromossômica acidental espontânea para $2n_1 + 2n_2$ em uma região do híbrido estéril, presumivelmente no tecido que finalmente se tornou uma flor e sofreu meiose para produzir gametas. No tecido $2n_1 + 2n_2$, existe um parceiro de pareamento para cada cromossomo e são produzidos gametas funcionais do tipo

$n_1 + n_2$. Esses gametas se fundem para produzir progênie alopoliploide $2n_1 + 2n_2$, que também é fértil. Esse tipo de alopoliploide por vezes é denominado **anfidiploide**, ou diploide duplo (Figura 17.7). O tratamento de um híbrido estéril com colchicina aumenta muito as chances de que os conjuntos cromossômicos sejam duplicados. Os anfidiploides atualmente são sintetizados de modo rotineiro dessa maneira. (Infelizmente para Karpechenko, seu anfidiploide apresentava as raízes de um repolho e as folhas de um rabanete.)

Quando o alopoliploide de Karpechenko foi cruzado com cada espécie parental — o repolho ou o rabanete — resultaram descendentes estéreis. A descendência do cruzamento com o repolho foi $2n_1 + n_2$, constituída a partir de um gameta $n_1 + n_2$ do alopoliploide e um gameta n_1 do repolho. Os cromossomos n_2 não apresentavam parceiros de pareamento; portanto, não poderia ocorrer uma meiose normal e a descendência era estéril. Assim, Karpechenko havia efetivamente criado uma nova espécie, sem possibilidade de troca gênica com o repolho ou com o rabanete. Ele denominou essa nova planta *Raphanobrassica*.

Na natureza, a alopoliploidia aparenta ter sido uma importante força na evolução de novas espécies de plantas. Um exemplo convincente é demonstrado pelo gênero *Brassica*, conforme ilustrado na Figura 17.8. Aqui, três espécies genitoras diferentes hibridizaram em todas as combinações de pares possíveis para formar novas espécies anfidiploides. A poliploidia natural já chegou a ser considerada algo raro, mas trabalhos recentes demonstraram que ela é um evento recorrente em muitas espécies de plantas. A utilização de marcadores de DNA tornou possível demonstrar que poliploides em qualquer população ou área que aparentam ser os mesmos são resultantes de muitas fusões anteriores independentes entre indivíduos geneticamente distintos das mesmas duas espécies parentais. Estima-se que 50% de todas as plantas angiospermas sejam poliploides, resultando da autopoliploidia ou da alopoliploidia. Como resultado de múltiplas poliploidizações, a quantidade de variação alélica em uma espécie poliploide é muito mais alta do que se acreditava anteriormente, talvez contribuindo para o seu potencial de adaptação.

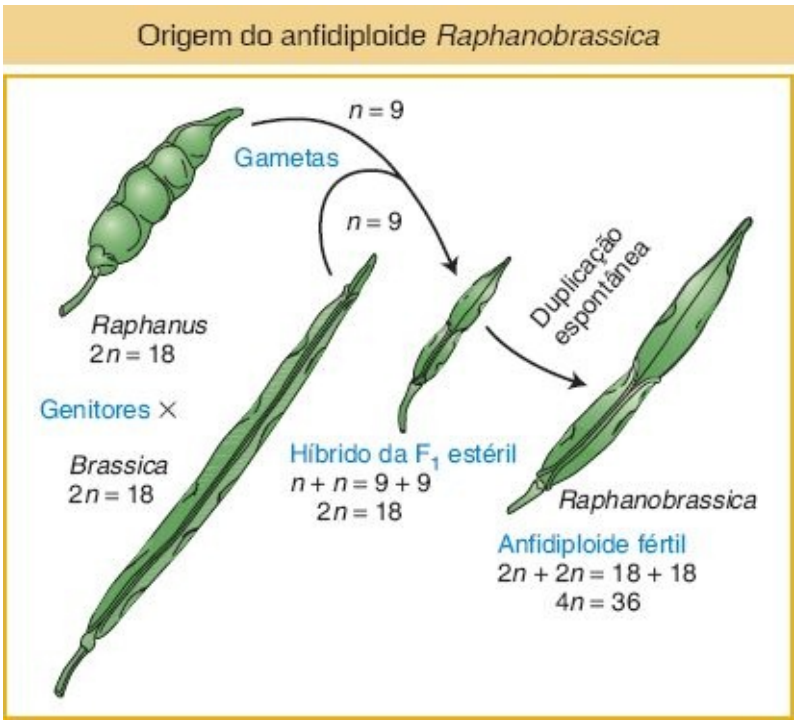


FIGURA 17.7 Na progênie de um cruzamento de repolho (*Brassica*) e rabanete (*Raphanus*), o anfidiploide fértil surgiu a partir da duplicação espontânea no híbrido estéril $2n = 18$.

Um alopoliploide natural particularmente interessante é o trigo do pão, *Triticum aestivum* ($6n = 42$). Ao estudar os seus parentes selvagens, geneticistas reconstruíram um provável histórico evolutivo dessa planta. A Figura 17.9 demonstra que o trigo é composto por dois conjuntos de três genomas ancestrais. Na meiose, o pareamento é sempre entre homólogos do mesmo genoma ancestral. Portanto, na meiose do trigo sempre existem 21 bivalentes.

As células da planta alopoliploide também podem ser produzidas artificialmente por meio da fusão de células diploides de espécies diferentes. Primeiramente, as paredes de duas células diploides são removidas por meio do tratamento com uma enzima e as membranas das duas células se fundem e se tornam uma. Com frequência os núcleos também se fundem, resultando no poliploide. Se a célula for nutrida com os hormônios e os nutrientes apropriados, ela se divide para se tornar uma pequena muda de planta alopoliploide, que em seguida pode ser transferida para o solo.

> CONCEITO-CHAVE Plantas alopoliploides podem ser sintetizadas por meio do cruzamento de espécies correlatas e da duplicação dos cromossomos do híbrido, ou por meio da fusão de células diploides.

Aplicações agrícolas. Variações no número de cromossomos têm sido exploradas para criar novas linhagens de plantas com características desejáveis. Seguem alguns exemplos.

Monoploides. A diploidia é um incômodo inerente para os agricultores. Quando eles desejam induzir e selecionar novas mutações recessivas que sejam favoráveis para fins agrícolas, as novas mutações não podem ser detectadas, a menos que sejam homozigotas. Os agricultores também podem desejar encontrar novas combinações favoráveis de alelos em diferentes *loci*, mas tais combinações favoráveis de alelos em heterozigotos serão desfeitas pela recombinação na meiose. Os monoploides proporcionam um modo de contornar alguns desses problemas.

Os monoploides podem ser artificialmente derivados dos produtos da meiose nas anteras de uma planta. Uma célula haploide destinada a tornar-se um grão de pólen pode, em vez disso, ser induzida por meio de tratamento a frio (sujeito a baixas temperaturas) para tornar-se um **embrioide**, uma pequena massa de células monoploides em divisão. O embrioide pode ser cultivado em ágar até a formação de uma muda monoploide, que em seguida pode ser plantada no solo e amadurecer (Figura 17.10).

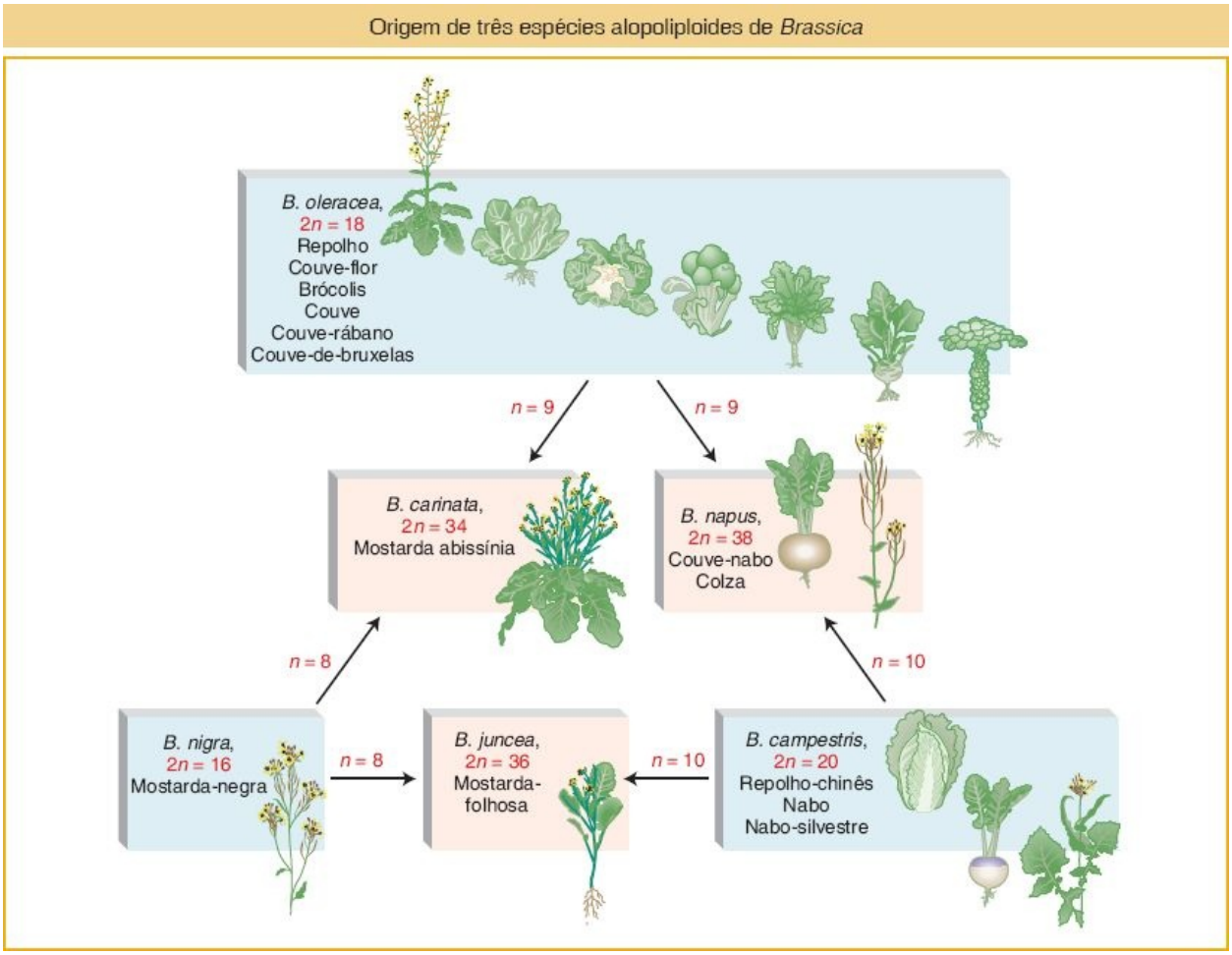


FIGURA 17.8 A alopoliploidia é importante na produção de novas espécies. No exemplo demonstrado, três espécies diploides de *Brassica* (*quadros verde-claros*) foram cruzadas em diferentes combinações para produzir seus alopoliploides (*quadros bege*). Alguns dos derivados agrícolas de algumas das espécies estão demonstrados dentro dos quadros.

As plantas monoploides podem ser exploradas de diversos modos. Em uma abordagem, elas são primeiramente examinadas em relação a combinações alélicas favoráveis que tenham surgido a partir da recombinação de alelos já presentes em um genitor diploide heterozigoto. Portanto, a partir de um genitor que é *A/a*; *B/b*, pode surgir uma combinação monoploide favorável *a*; *b*. Em seguida, o monoploide pode ser submetido à duplicação cromossômica até produzir células diploides homozigotas, *a/a*; *b/b*, que conseguem realizar reprodução normal.

Outra abordagem é tratar as células monoploides basicamente como uma

população de organismos haploides em um procedimento de mutagênese e seleção. Uma população de células monoploides é isolada, suas paredes são removidas por meio de tratamento enzimático e elas são expostas a um mutágeno. Em seguida elas são plaqueadas em um meio seletivo para algum fenótipo desejável. Essa abordagem tem sido utilizada para a seleção em relação à resistência a compostos tóxicos produzidos por um parasita de plantas, bem como para a seleção em relação à resistência a herbicidas que estão sendo utilizados por fazendeiros para matar pragas nas plantações. As mudas resistentes finalmente crescem em plantas monoploides, cujo número cromossômico em seguida pode ser duplicado com a utilização de colchicina, levando a um diploide homozigoto resistente. Essas técnicas poderosas podem contornar o processo normalmente lento do cultivo de plantas com base na meiose. Elas têm sido aplicadas com sucesso em importantes plantações, tais como soja e tabaco.

> CONCEITO-CHAVE Geneticistas podem criar novas linhagens de plantas por meio da produção de monoploides com genótipos favoráveis e, em seguida, duplicar seus cromossomos até formar diploides homozigotos férteis.

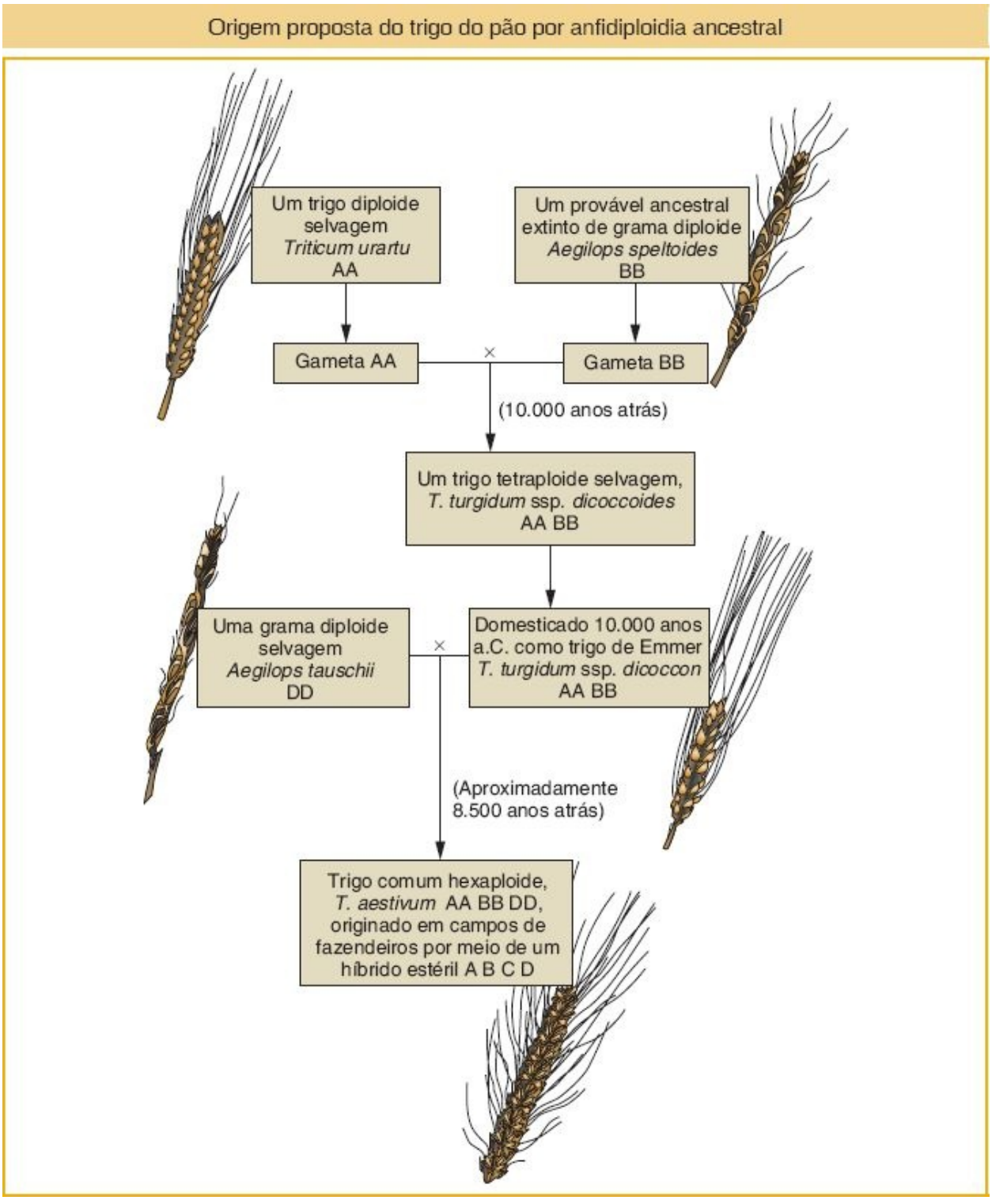


FIGURA 17.9 O trigo moderno teve origem a partir de dois casos ancestrais de anfidiploidia, primeiramente por meio de gametas não reduzidos, depois por meio de um intermediário estéril.

Autotriploides. As bananas que se encontram disponíveis comercialmente de

modo amplo são triploides estéreis com 11 cromossomos em cada conjunto (3*n* = 33). A expressão mais óbvia da esterilidade das bananas é a ausência de sementes na fruta que ingerimos. (As manchas pretas nas bananas não são sementes; as sementes de banana são muito duras — podem quebrar os dentes.) Melancias sem sementes são outro exemplo da exploração comercial da triploidia em plantas.

Autotetraploides. Muitas plantas autotetraploides foram desenvolvidas como cultivos comerciais para se obter uma vantagem de seu tamanho aumentado (Figura 17.11). Frutos e flores grandes são particularmente favorecidos.

Alopoliploides. A alopoliploidia (formação de poliploides entre espécies diferentes) foi importante na produção de plantações modernas. O algodão do Novo Mundo é um alopoliploide natural que surgiu espontaneamente, assim como o trigo. Os alopoliploides também são sintetizados artificialmente para combinar as características úteis das espécies parentais em um tipo. Apenas um anfidiploide sintético chegou a ser utilizado comercialmente de modo amplo, um cultivo conhecido como *Triticale*. Ele é um anfidiploide entre o trigo (*Triticum*, 6*n* = 42) e o centeio (*Secale*, 2*n* = 14). Portanto, em relação ao *Triticale*, 2*n* = 2 × (21 + 7) = 56. Essa nova planta combina a alta produtividade do trigo com a resistência do centeio.

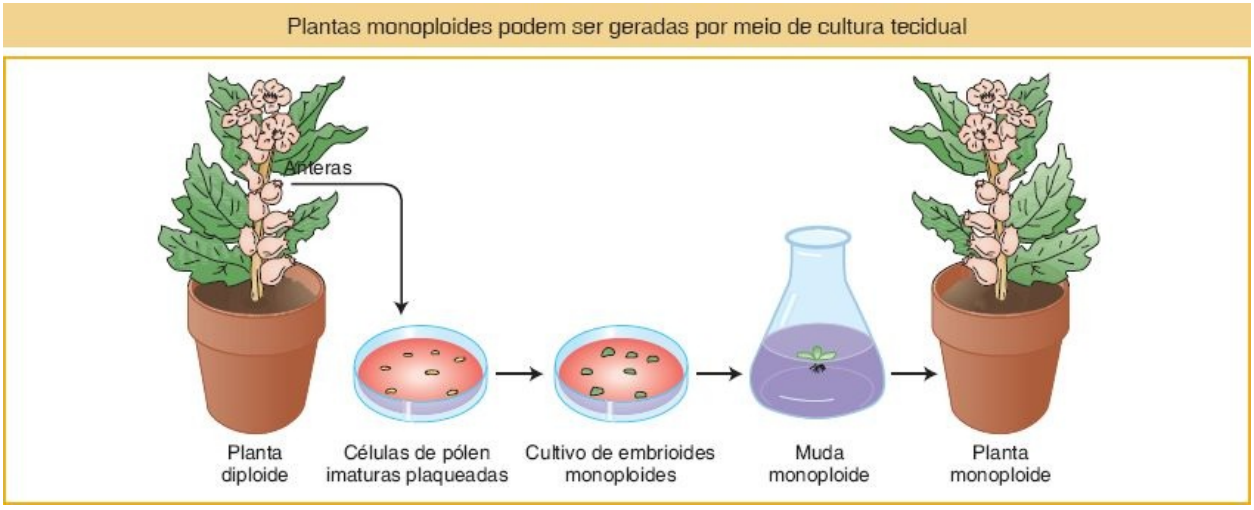


FIGURA 17.10 Plantas monoploides podem ser derivadas artificialmente de células destinadas a tornar-se grãos de pólen por meio da exposição das células ao tratamento a frio em cultura de tecido.

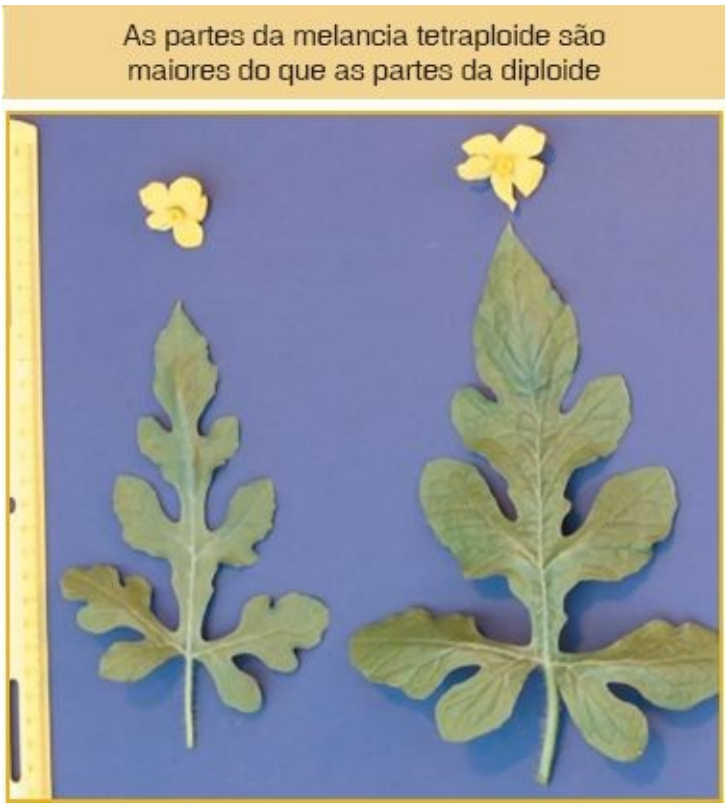


FIGURA 17.11 Folhas e flores de melancia diploide (*esquerda*) e tetraploide (*direita*). (Michael E. Compton, University of Wisconsin — Platteville.)

Animais poliploides. Conforme observado anteriormente, a poliploidia é mais comum em plantas do que em animais, mas existem casos de animais poliploides de ocorrência natural. Espécies poliploides de platelmintos, sanguessugas e camarões de água salgada se reproduzem por meio de partenogênese. *Drosophilae* triploides e tetraploides foram sintetizadas experimentalmente. Entretanto, os exemplos não estão limitados a essas chamadas formas inferiores. Anfíbios e répteis poliploides de ocorrência natural são surpreendentemente comuns. Eles apresentam diversos modos de reprodução: espécies poliploides de rãs e sapos têm na reprodução sexuada, enquanto salamandras e lagartos poliploides são partenogenéticos. Salmonidae (a família de peixes que inclui o salmão e a truta) é um exemplo conhecido das diversas espécies de animais que aparentam ter sido originadas por poliploidia ancestral.

A esterilidade dos triploides tem sido explorada comercialmente em animais, bem como em plantas. Ostras triploides têm sido desenvolvidas em virtude de

apresentarem uma vantagem comercial sobre as diploides. Os diploides passam por uma temporada de produção de ovos, quando não são palatáveis, mas os triploides estéreis não põem ovos e são palatáveis durante o ano todo.

Aneuploidia

A aneuploidia é a segunda maior categoria de aberrações cromossômicas na qual o número de cromossomos é anormal. Um aneuploide é um organismo cujo número de cromossomos difere daquele do tipo selvagem em parte de um conjunto cromossômico. Em geral, o conjunto cromossômico aneuploide difere do tipo selvagem em apenas um cromossomo, ou em um pequeno número de cromossomos. Um aneuploide pode apresentar um número de cromossomos superior ou inferior àquele do tipo selvagem. A nomenclatura aneuploide (ver Tabela 17.1) tem por base o número de cópias do cromossomo específico no estado aneuploide. Em relação aos autossomos em organismos diploides, o aneuploide $2n + 1$ é **trissômico**, $2n - 1$ é **monossômico** e $2n - 2$ (o "— 2" representa a perda de ambos os homólogos de um cromossomo) é **nulissômico**. Em haploides, $n + 1$ é **dissômico**. É utilizada uma anotação especial para descrever os cromossomos sexuais aneuploides, tendo em vista que ela deve lidar com dois cromossomos diferentes. A notação meramente lista as cópias de cada cromossomo sexual, tal como XXY, XYY, XXX ou XO (o "O" refere-se à ausência de um cromossomo e é incluído para demonstrar que o símbolo X único não é um erro tipográfico).

Não disjunção. A causa da maior parte das aneuploidias é a **não disjunção** no curso da meiose ou da mitose. *Disjunção* é outra palavra para a segregação normal de cromossomos ou cromátides homólogas para polos opostos nas divisões meióticas ou mitóticas. A *não disjunção* é uma falha desse processo, na qual dois cromossomos ou duas cromátides se dirigem incorretamente para um polo e não para o outro.

A não disjunção *mitótica* pode ocorrer à medida que as células se dividem durante o desenvolvimento. Como resultado, partes do corpo serão aneuploides (*setores* aneuploides). A não disjunção *meiótica* é observada mais comumente.

Nesse caso, os produtos da meiose são aneuploides, levando a descendentes totalmente aneuploides. Na não disjunção meiótica, os cromossomos podem falhar em se separar na primeira ou na segunda divisão meiótica (Figura 17.12). De qualquer modo, são produzidos gametas n — 1 e n + 1. Se um gameta n — 1 é fertilizado por um gameta n, é produzido um zigoto monossômico ($2n$ — 1). A fusão de um gameta n + 1 e um gameta n produz um trissômico $2n$ + 1.

> CONCEITO-CHAVE Os organismos aneuploides resultam principalmente de não disjunção na meiose parental.

A não disjunção meiótica gera produtos aneuploides

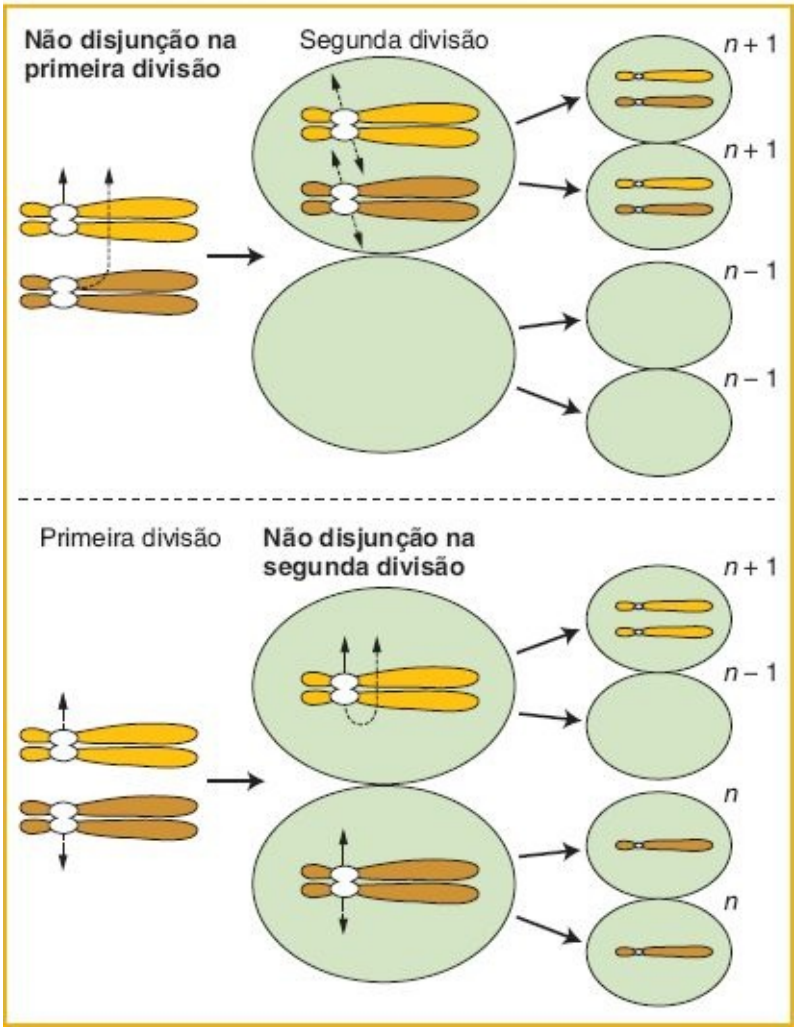


FIGURA 17.12 Produtos aneuploides da meiose (ou seja, os gametas) são produzidos por meio da não disjunção na primeira ou na segunda divisão meiótica. Observe que todos os outros cromossomos estão presentes em números normais, incluindo nas células nas quais nenhum cromossomo está demonstrado.

A não disjunção ocorre espontaneamente. Assim como a maior parte das mutações gênicas, ela é um exemplo de uma falha ao acaso de um processo celular básico. Os processos moleculares precisos que falham não são conhecidos, mas, em sistemas experimentais, a frequência de não disjunção pode ser aumentada por meio da interferência com a polimerização dos microtúbulos, inibindo, assim, a movimentação cromossômica normal. Aparentemente, a disjunção apresenta maior probabilidade de erro na meiose I. Essa falha não é uma surpresa, tendo em vista que a disjunção normal na anáfase I requer que as cromátides homólogas da tétrade permaneçam pareadas durante a prófase I e a metáfase I, além de *crossovers*. Contrariamente, a disjunção adequada na anáfase II ou na mitose requer que o centrômero se divida adequadamente, mas não requer o pareamento cromossômico ou o *crossing over*.

Os *crossovers* são um componente necessário do processo de disjunção normal. De algum modo, a formação de um quiasma ajuda a manter um bivalente unido e assegura que as duas díades irão se dirigir para os polos opostos. Na maior parte dos organismos, *crossing over* é suficiente para assegurar que todos os bivalentes apresentarão no mínimo um quiasma por meiose. Em *Drosophila*, muitos dos cromossomos que sofreram não disjunção observados em gametas dissômicos ($n + 1$) são não recombinantes, demonstrando que eles surgem a partir de meioses nas quais não ocorre um *crossing over* naquele cromossomo. Foram realizadas observações semelhantes em trissomias humanas. Além disso, em inúmeros organismos experimentais diferentes, as mutações que interferem com a recombinação apresentam o efeito de aumentar maciçamente a frequência de não disjunção na meiose I. Todas essas observações fornecem evidências em relação ao papel do *crossing over* na manutenção do pareamento cromossômico; na ausência dessas associações, os cromossomos são vulneráveis à não disjunção na anáfase I.

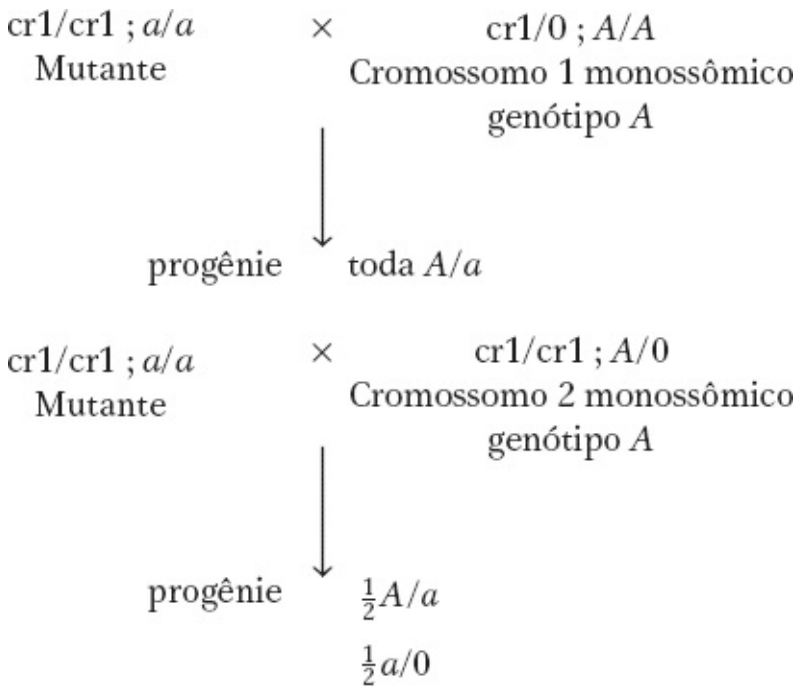
> CONCEITO-CHAVE Os *crossovers* são necessários para manter os bivalentes pareados até a anáfase I. Se o *crossing over* por algum motivo falha, ocorre a não disjunção de primeira divisão.

Monossômicos (2*n* — 1). Os monossômicos apresentam a ausência de uma cópia de um cromossomo. Na maior parte dos organismos diploides, a ausência de uma cópia de um cromossomo de um par é deletéria. Em seres humanos, os monossômicos em relação a qualquer dos autossomos morrem *in utero*. Muitos monossômicos do cromossomo X também morrem *in utero*, mas alguns são viáveis. Um complemento cromossômico humano de 44 autossomos mais um único X produz uma condição conhecida como **síndrome de Turner**, representada como XO. As pessoas afetadas apresentam um fenótipo característico: elas são mulheres estéreis, de estatura baixa e com frequência apresentam uma frouxidão da pele que se estende entre o pescoço e os ombros (Figura 17.13). Embora a sua inteligência esteja próxima do normal, algumas de suas funções cognitivas específicas são defeituosas. Aproximadamente 1 em 5.000 nascimentos do sexo feminino demonstra a síndrome de Turner.

Geneticistas utilizaram plantas monossômicas viáveis para mapear alelos mutantes recessivos recém-descobertos em um cromossomo específico. Por exemplo, pode-se produzir um conjunto de linhagens monossômicas, cada uma sabidamente sem um cromossomo diferente. Os homozigotos em relação ao novo alelo mutante são cruzados com cada linhagem monossômica e a progênie de cada cruzamento é inspecionada em relação ao fenótipo recessivo. O aparecimento do fenótipo recessivo identifica o cromossomo que não apresenta uma cópia como aquele no qual o gene normalmente está localizado. O teste funciona porque metade dos gametas de um monossômico 2*n* — 1 fértil será *n* — 1 e, quando um gameta *n* — 1 é fertilizado por um gameta que contém uma nova mutação no cromossomo homólogo, o alelo mutante será o único alelo daquele gene presente e, portanto, será expresso.

Para ilustrar, presumiremos que um gene *A/a* se encontra no cromossomo 2. Prevê-se que cruzamentos de *a/a* com monossômicos para o cromossomo 1 e o

cromossomo 2 produzam resultados diferentes (o cromossomo 1 é abreviado cr1):



Trissômicos (2*n* + 1). Os trissômicos contêm uma cópia extra de um cromossomo. Em organismos diploides em geral, o desequilíbrio cromossômico da condição trissômica pode resultar em anormalidade ou morte. Entretanto, existem muitos exemplos de trissômicos viáveis. Além disso, os trissômicos podem ser férteis. Quando as células de alguns organismos trissômicos são observadas ao microscópio no momento do pareamento cromossômico meiótico, observa-se que os cromossomos trissômicos formam um grupo associado de três (um trivalente), enquanto os outros cromossomos formam bivalentes regulares.

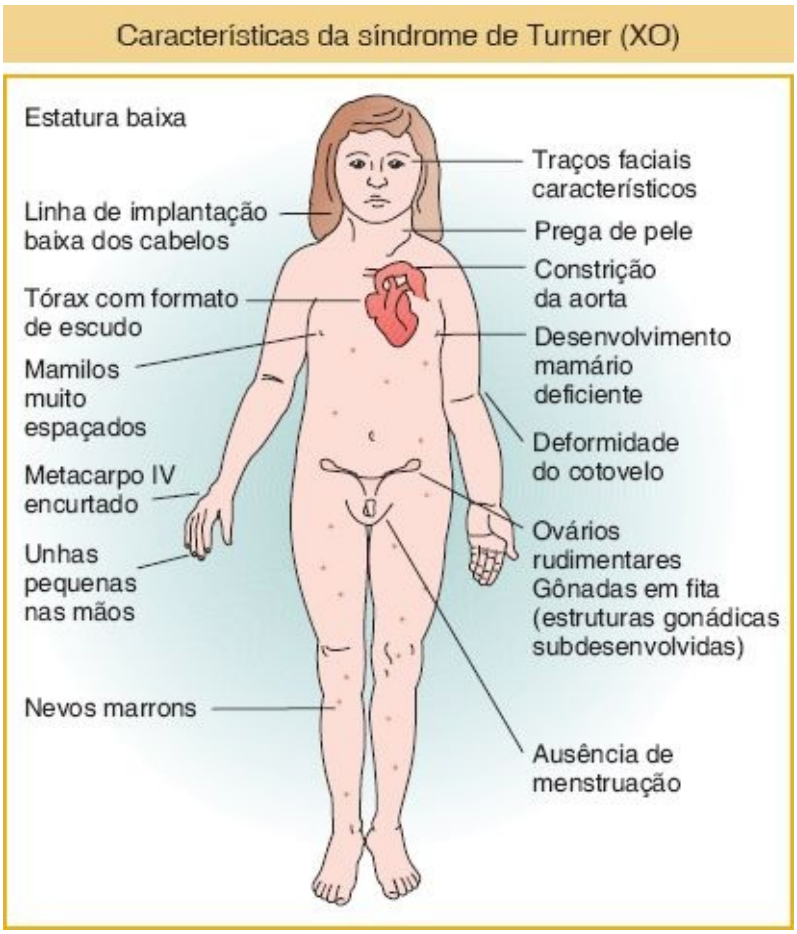


FIGURA 17.13 A síndrome de Turner resulta da presença de um único cromossomo X (XO).

Quais razões genéticas podem ser esperadas para genes no cromossomo trissômico? Consideraremos um gene *A* que está próximo do centrômero naquele cromossomo e assumiremos que o genótipo é *A/a/a*. Além disso, postularemos que, na anáfase I, os dois centrômeros pareados no trivalente passam para polos opostos e que o outro centrômero passa aleatoriamente para qualquer polo. Em seguida, podemos prever as três segregações igualmente frequentes demonstradas na Figura 17.14. Essas segregações resultam em uma razão gamética geral, conforme demonstrado nos seis compartimentos da Figura 17.14; ou seja:

$$\frac{1}{6}\ A$$

$$\frac{2}{6}\ a$$

$$\frac{2}{6}\ A/a$$

$\frac{1}{6}$ *a/a*

Se estiver disponível um conjunto de linhagens, cada uma carreando um cromossomo trissômico diferente, então uma mutação gênica pode ser localizada em um cromossomo por meio da determinação de qual das linhagens proporciona uma razão trissômica do tipo precedente.

Existem diversos exemplos de trissomias humanas viáveis. Diversos tipos de trissômicos de cromossomos sexuais podem viver até a fase adulta. Cada um desses tipos é observado a uma frequência de aproximadamente 1 em 1.000 nascimentos vivos do sexo relevante. (Ao considerar as trissomias dos cromossomos sexuais humanos, relembre que o sexo dos mamíferos é determinado pela presença ou pela ausência do cromossomo Y.) A combinação XXY resulta na **síndrome de Klinefelter**. Pessoas com essa síndrome são homens que apresentam constituições magras, um QI levemente comprometido e são estéreis (Figura 17.15). Outra combinação anormal, XYY, apresenta um histórico controverso. Foram realizadas tentativas de ligar a condição XYY a uma predisposição à violência. Entretanto, atualmente está claro que uma condição XYY de modo algum garante tal comportamento. A maior parte dos homens XYY é fértil. As meioses demonstram pareamento normal do X com um dos Y; o outro Y não pareia e não é transmitido para os gametas. Portanto, os gametas contêm X ou Y, nunca YY ou XY. Trissômicos triplo X (XXX) são mulheres fenotipicamente normais e férteis. A meiose demonstra o pareamento de apenas dois cromossomos X; o terceiro não pareia. Portanto, os ovócitos apresentam apenas um X e, assim como aquela de homens XYY, a condição não é transmitida para a progênie.

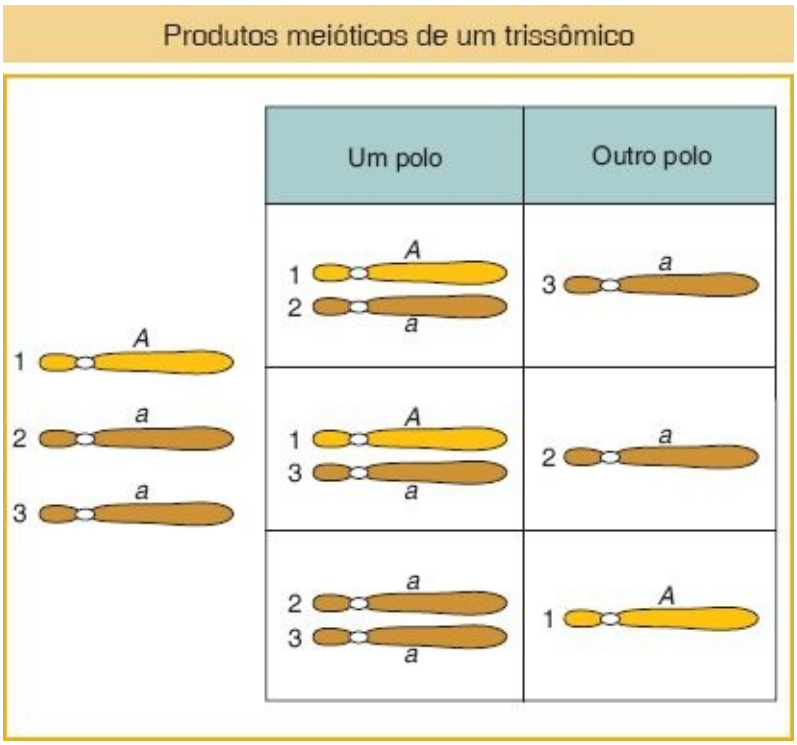


FIGURA 17.14 Três segregações igualmente prováveis podem ocorrer na meiose de um trissômico *A/a/a*, produzindo os genótipos demonstrados.

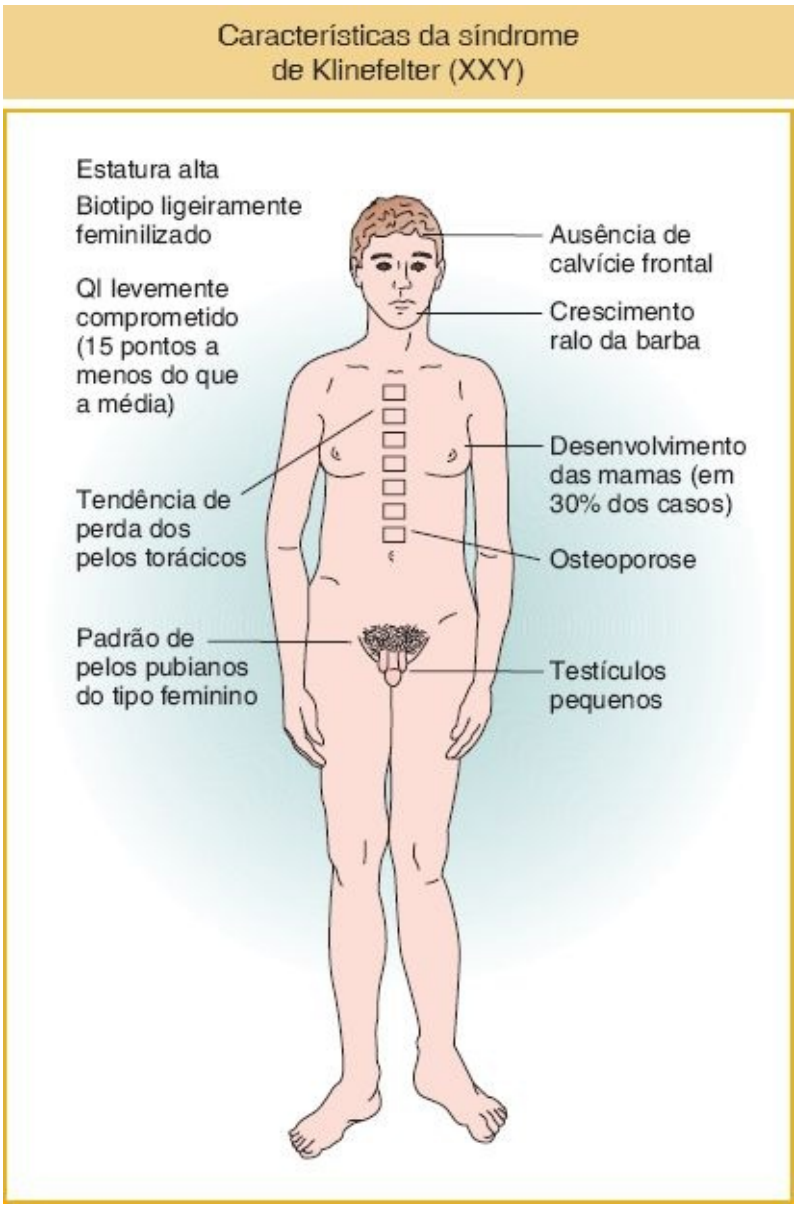


FIGURA 17.15 A síndrome de Klinefelter resulta da presença de dois cromossomos X e um cromossomo Y.

Das trissomias humanas, o tipo mais familiar é a **síndrome de Down** (Figura 17.16), discutida brevemente no início do capítulo. A frequência da síndrome de Down é de aproximadamente 0,15% de todos os nascimentos vivos. A maioria das pessoas afetadas apresenta uma cópia extra do cromossomo 21, causada pela não disjunção do cromossomo 21 em um genitor cromossomicamente normal. Nesse tipo *esporádico* de síndrome de Down, não existe um histórico familiar de aneuploidia. Alguns tipos mais raros de síndrome de Down surgem a partir de

translocações (um tipo de rearranjo cromossômico, discutido posteriormente no capítulo); nesses casos, conforme veremos, a síndrome de Down recorre no heredograma, tendo em vista que a translocação pode ser transmitida do genitor para o filho.

Os fenótipos combinados que compõem a síndrome de Down incluem retardo mental (com QI na faixa de 20 a 50); face achatada e larga; olhos com dobra epicântica; estatura baixa; mãos curtas com uma prega na parte intermediária; e uma língua grande e sulcada. As mulheres podem ser férteis e produzir progênie normal ou trissômica, mas os homens são estéreis, com raríssimas exceções. A expectativa de vida média é de aproximadamente 17 anos e apenas 8% das pessoas com síndrome de Down sobrevivem até depois dos 40 anos de idade.

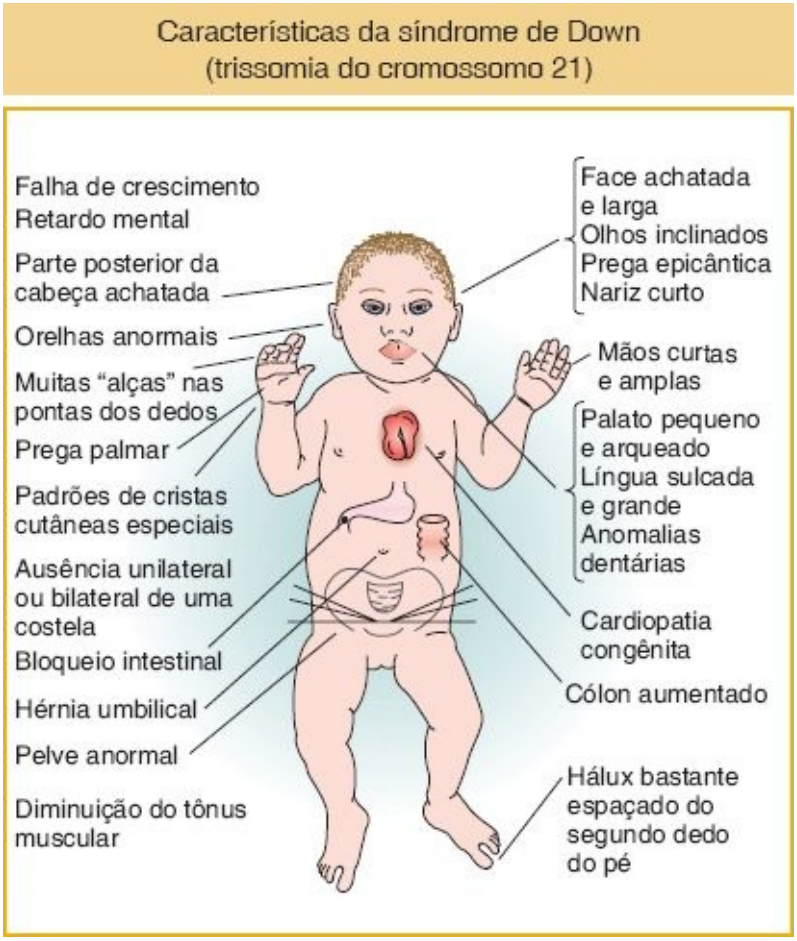


FIGURA 17.16 A síndrome de Down resulta da presença de uma cópia extra do cromossomo 21.

A incidência da síndrome de Down está relacionada com a idade materna: mães mais velhas apresentam um risco muito elevado de ter um filho com síndrome de Down (Figura 17.17). Por esse motivo, a análise cromossômica do feto (por meio de amniocentese ou de amostra de vilosidades coriônicas) atualmente é recomendada para mães gestantes mais velhas. Também foi demonstrado um efeito menos pronunciado da idade paterna.

Embora o efeito da idade materna tenha sido conhecido há muitos anos, ainda não se sabe a sua causa. Não obstante, existem algumas correlações biológicas interessantes. Com a idade, possivelmente o cromossomo bivalente apresenta menos probabilidade de permanecer unido durante a prófase I da meiose. A parada meiótica dos ovócitos (meiócitos femininos) no final da prófase I é um fenômeno comum em muitos animais. Em mulheres, todos os ovócitos param no diplóteno antes do nascimento. A meiose é retomada a cada período menstrual, o que significa que os cromossomos bivalentes devem permanecer adequadamente associados por até cinco ou mais décadas. Se especularmos que essas associações apresentam uma probabilidade crescente de ruptura por acidente na medida em que o tempo passa, podemos imaginar um mecanismo que contribui para o aumento da não disjunção materna com a idade. Consistente com essa especulação, a maior parte das não disjunções relacionadas com o efeito da idade materna ocorre em virtude da não disjunção na anáfase I, não na anáfase II.

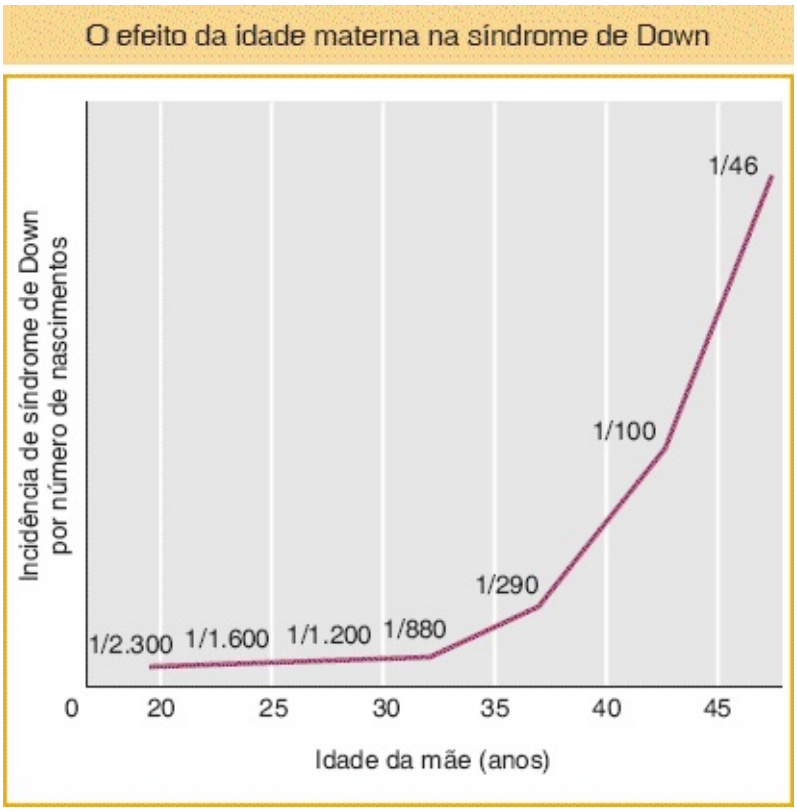


FIGURA 17.17 Mães mais velhas apresentam uma proporção mais alta de bebês com síndrome de Down do que mães mais jovens. (Dados de L. S. Penrose e G. F. Smith, *Down's Anomaly*. Little, Brown and Company, 1966.)

Os únicos outros trissômicos autossômicos humanos que sobrevivem até o nascimento são aqueles com a trissomia do cromossomo 13 (síndrome de Patau) e a trissomia do cromossomo 18 (síndrome de Edwards). Ambos apresentam anormalidades físicas e mentais graves. A síndrome fenotípica da trissomia do cromossomo 13 inclui fenda labial; cabeça pequena e malformada; pés "arqueados"; e expectativa de vida média de 130 dias. A trissomia do cromossomo 18 inclui orelhas tipo fauno (implantação baixa), mandíbula estreita, pelve estreita e pés arqueados; quase todos os bebês com trissomia do 18 morrem nas primeiras semanas após o nascimento. Todos os outros trissômicos morrem *in utero*.

Conceito de balanço gênico

Ao considerar a euploidia aberrante, observamos que um aumento no número de

conjuntos de cromossomos completos está correlacionado com o maior tamanho do organismo, mas que a forma e as proporções gerais do organismo permanecem em grande parte as mesmas. Contrariamente, a aneuploidia autossômica tipicamente altera a forma e as proporções do organismo de modos característicos.

As plantas tendem a ser um pouco mais tolerantes à aneuploidia do que os animais. Estudos no estramônio (*Datura stramonium*) fornecem um exemplo clássico dos efeitos da aneuploidia e da poliploidia. Nessa planta, o número de cromossomos haploides é 12. Conforme esperado, o estramônio poliploide é proporcional, assim como o diploide normal, apenas maior. Contrariamente, cada um dos 12 trissômicos possíveis é desproporcional, mas de diferentes modos entre si, conforme exemplificado pelas alterações na forma da cápsula da semente (Figura 17.18). As 12 trissomias diferentes levam a 12 alterações diferentes e características na forma da cápsula. De fato, essas características e outras dos trissômicos individuais são tão confiáveis que a síndrome fenotípica pode ser utilizada para identificar plantas que carreiam uma trissomia em particular. De modo semelhante, os 12 monossômicos são, eles próprios, diferentes entre si e de cada um dos trissômicos. Em geral, um monossômico em relação a um cromossomo em particular é mais gravemente anormal do que o trissômico correspondente.

Observamos tendências semelhantes em animais aneuploides. Na mosca-das-frutas *Drosophila*, os únicos aneuploides autossômicos que sobrevivem até a fase adulta são os trissômicos e os monossômicos em relação ao cromossomo 4, que é o menor cromossomo de *Drosophila*, representando apenas aproximadamente 1 a 2% do genoma. Os trissômicos em relação ao cromossomo 4 são apenas levemente afetados e são muito menos anormais do que os monossômicos em relação ao cromossomo 4. Em seres humanos, nenhum monossômico autossômico sobrevive até o nascimento, mas, conforme já declarado, três tipos de trissômicos autossômicos podem sobreviver. Assim como é verdadeiro para o estramônio aneuploide, cada um desses três trissômicos demonstra síndromes fenotípicas únicas em virtude dos efeitos especiais de doses alteradas de cada um desses cromossomos.

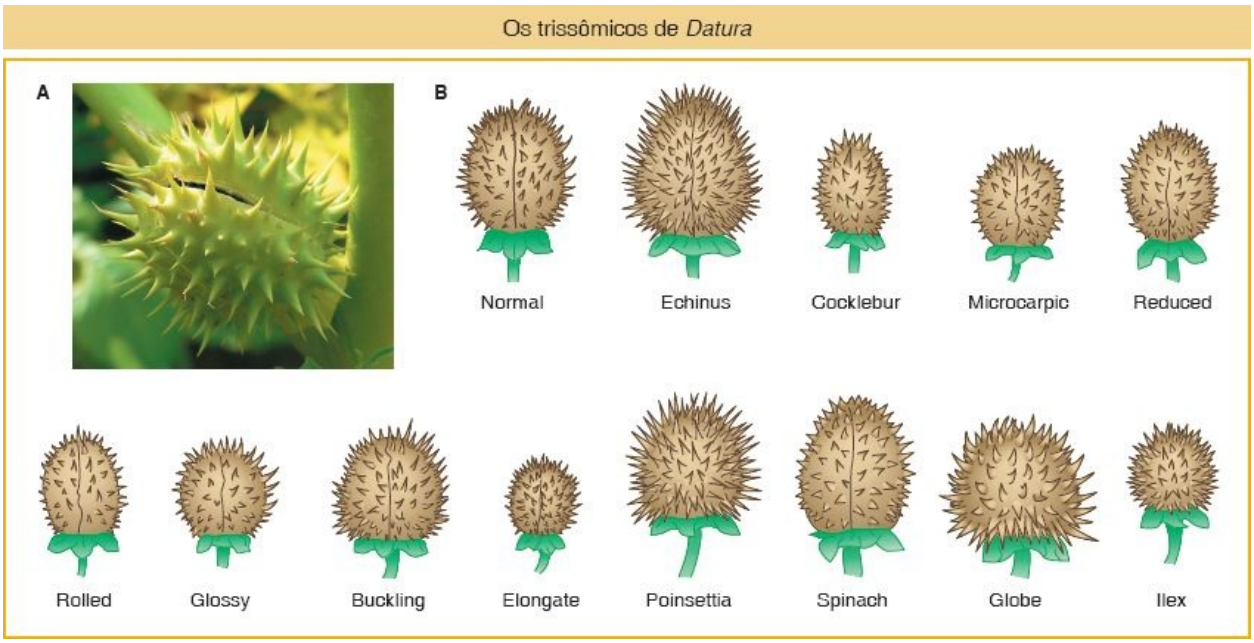


FIGURA 17.18 Cada um dos 12 trissômicos possíveis de *Datura* é desproporcional de um modo diferente. **A.** Fruto de *Datura*. **B.** Cada desenho é do fruto de um trissômico diferente, com sua respectiva denominação. (*A.* iStockphoto/Thinkstock.)

Por que os aneuploides são muito mais anormais do que os poliploides? Por que a aneuploidia em relação a cada cromossomo apresenta seus próprios efeitos fenotípicos característicos? E por que os monossômicos são em geral mais gravemente afetados do que os trissômicos correspondentes? As respostas aparentam ser certamente uma questão de **balanço gênico**. Em um euploide, a razão de genes em qualquer cromossomo com relação aos genes em outros cromossomos é sempre 1:1, independentemente de estarmos considerando um monoploide, diploide, triploide ou tetraploide. Por exemplo, em um tetraploide, em razão ao gene *A* no cromossomo 1 e ao gene *B* no cromossomo 2, a razão é de 4 *A*:4 *B*, ou 1:1. Contrariamente, em um aneuploide, a razão de genes no cromossomo aneuploide em relação aos genes em outros cromossomos difere do tipo selvagem em 50%:50% em relação aos monossômicos; 150% para os trissômicos. Com a utilização do mesmo exemplo anterior, em um trissômico em razão ao cromossomo 2, observamos que a razão dos genes *A* e *B* é de 2 *A*:3 *B*. Portanto, podemos verificar que os genes aneuploides estão desbalanceados. Como o fato de estarem desbalanceados nos ajuda a responder as questões

levantadas?

Em geral, a quantidade de transcritos produzida por um gene é diretamente proporcional ao número de cópias daquele gene em uma célula. Ou seja, em relação a um determinado gene, a taxa de transcrição está diretamente relacionada com o número de moldes de DNA disponíveis. Portanto, quanto mais cópias do gene, mais transcritos são produzidos e mais do produto proteico correspondente é produzido. Essa relação entre o número de cópias de um gene e a quantidade do produto gênico produzida é denominada **efeito de dosagem gênica**.

Podemos inferir que a fisiologia normal em uma célula depende da razão adequada de produtos gênicos na célula euploide. Essa razão é o balanço gênico normal. Se a dosagem relativa de determinados genes for alterada — por exemplo, em virtude da remoção de uma das duas cópias de um cromossomo (ou até mesmo de um segmento dele) —, podem surgir desequilíbrios fisiológicos nas vias celulares.

Em alguns casos, os desbalanceamentos de aneuploidia resultam dos efeitos de alguns poucos genes "importantes", cuja dosagem foi alterada, em vez de alterações na dosagem de todos os genes em um cromossomo. Tais genes podem ser considerados *haploanormais* (resultando em um fenótipo anormal, se presente apenas uma vez), ou *triploanormais* (resultando em um fenótipo anormal, se presente em três cópias), ou ambos. Eles contribuem significativamente para as síndromes fenotípicas aneuploides. Por exemplo, o estudo de pessoas trissômicas em relação a apenas uma parte do cromossomo 21 tornou possível localizar os genes que contribuem para a síndrome de Down em diversas regiões do cromossomo 21; os resultados indicam que alguns aspectos do fenótipo podem ocorrer em virtude da tripla anormalidade em relação a genes importantes únicos nessas regiões cromossômicas. Além dos efeitos desses genes importantes, outros aspectos das síndromes aneuploides provavelmente resultam dos efeitos cumulativos da aneuploidia de diversos genes cujos produtos estão todos desbalanceados. Indubitavelmente, o fenótipo aneuploide inteiro resulta de uma combinação dos efeitos do desbalanço de alguns genes importantes, juntamente com um desbalanço cumulativo de muitos genes menos importantes.

Entretanto, o conceito de balanço gênico não nos informa o motivo pelo qual a

apresentação de poucos produtos gênicos (monossomia) é muito pior para um organismo do que a apresentação de muitos produtos gênicos (trissomia). Paralelamente, podemos indagar por que existem muito mais genes haploanormais do que triploanormais. Uma chave para explicar a anormalidade extrema dos monossômicos é que quaisquer alelos recessivos deletérios presentes em um autossomo monossômico serão automaticamente expressos.

Como aplicamos a ideia de balanço gênico para casos de aneuploidia de cromossomos sexuais? O balanço gênico é mantido também em relação aos cromossomos sexuais, mas também devemos levar em consideração as propriedades especiais dos cromossomos sexuais. Em organismos com determinação sexual XY, o cromossomo Y aparenta ser um cromossomo X degenerado, no qual existem muito poucos genes funcionais além de alguns relacionados com a própria determinação sexual, à produção de espermatozoides, ou ambas. O cromossomo X, por outro lado, contém muitos genes relacionados com processos celulares básicos ("genes de manutenção"), presentes no cromossomo que por fim evoluiu para o cromossomo X. Os mecanismos de determinação sexual XY provavelmente evoluíram independentemente de 10 a 20 vezes em diferentes grupos taxonômicos. Por exemplo, aparentemente existe um mecanismo de determinação sexual para todos os mamíferos, mas ele é completamente diferente do mecanismo que regula a determinação sexual XY nas moscas-das-frutas.

Em um sentido, os cromossomos X são naturalmente aneuploides. Em espécies com um sistema de determinação sexual XY, as fêmeas apresentam dois cromossomos X, enquanto os machos apresentam apenas um. Não obstante, os genes de manutenção do cromossomo X são expressos de modo aproximadamente igual por célula em fêmeas e em machos. Em outras palavras, existe **compensação de dose**. Como essa compensação é conquistada? A resposta depende do organismo. Nas moscas-das-frutas, o cromossomo X do macho aparenta estar hiperativado, possibilitando que ele seja transcrito no dobro da taxa de cada cromossomo X na fêmea. Como resultado, o macho XY de *Drosophila* apresenta uma dosagem do gene X equivalente àquela de uma fêmea XX. Em mamíferos, contrariamente, a regra é que não importa quantos

cromossomos X estejam presentes, existe apenas um cromossomo X ativo em relação à transcrição em cada célula somática. Essa regra proporciona à fêmea XX de mamíferos uma dosagem gênica do X equivalente àquela de um macho XY. A compensação de dose em mamíferos é conquistada por meio da inativação do cromossomo X. Uma fêmea com dois cromossomos X, por exemplo, é um mosaico de dois tipos celulares, no qual um ou outro cromossomo X está ativo. Examinamos esse fenômeno no Capítulo 12. Portanto, indivíduos XY e XX produzem as mesmas quantidades de produtos dos genes de manutenção do cromossomo X. A inativação do cromossomo X também explica o motivo pelo qual os seres humanos triplo X são fenotipicamente normais: apenas um dos três cromossomos X está ativo em relação à transcrição em uma determinada célula. De modo semelhante, um homem XXY é apenas moderadamente afetado, tendo em vista que apenas um de seus dois cromossomos X está ativo em cada célula.

Por que os indivíduos XXY são anormais, tendo em vista que indivíduos triplo X são fenotipicamente normais? Ocorre que alguns poucos genes dispersos em um "X inativo" ainda estão ativos em relação à transcrição. Em homens XXY, esses genes são transcritos no dobro do nível em que são transcritos em homens XY. Em mulheres XXX, por outro lado, os poucos genes transcritos estão ativos em apenas 1,5 vez o nível em que são transcritos em mulheres XX. Esse nível mais baixo de "aneuploidia funcional" em XXX do que em XXY, somado ao fato de que os genes do X ativo aparentam levar à feminilização, podem explicar o fenótipo feminilizado dos machos XXY. A gravidade da síndrome de Turner (XO) pode ocorrer em virtude dos efeitos deletérios da monossomia e a mais baixa atividade dos genes transcritos das fêmeas X (em comparação às XX). Conforme normalmente é observado em relação aos aneuploides, a monossomia do cromossomo X produz um fenótipo mais anormal do que a presença de uma cópia extra do mesmo cromossomo (mulheres triplo X ou homens XXY).

A dosagem gênica também é importante nos fenótipos de poliploides. Zigotos poliploides humanos surgem por meio de diversos tipos de erros na divisão celular. A maior parte morre *in utero*. Ocasionalmente, nascem bebês triploides, mas nenhum sobrevive. Esse fato aparenta violar o princípio de que os poliploides são mais normais do que os aneuploides. A explicação para essa

contradição parece estar relacionada com a compensação de dose do cromossomo X. Parte da regra em relação ao balanço gênico em organismos que apresentam um único X ativo parece ser que deve haver um X ativo para cada duas cópias do complemento cromossômico autossômico. Portanto, observa-se que algumas células em mamíferos triploides apresentam um X ativo, enquanto outras, surpreendentemente, apresentam dois. Nenhuma situação está em balanço com os genes autossômicos.

> CONCEITO-CHAVE A aneuploidia é quase sempre deletéria em virtude do desbalanceamento gênico: a razão de genes é diferente daquela em euploides e essa diferença interfere na função normal do genoma.

17.2 Alterações na estrutura dos cromossomos

As alterações na estrutura dos cromossomos, denominadas **rearranjos**, envolvem diversas classes de eventos importantes. Um segmento cromossômico pode ser perdido, constituindo uma **deleção**, ou duplicado, para formar uma **duplicação**. A orientação de um segmento dentro do cromossomo pode ser revertida, constituindo uma **inversão**. Ou um segmento pode ser movido para um cromossomo diferente, constituindo uma **translocação**. A quebra do DNA é uma causa importante de cada um desses eventos. Ambos os filamentos de DNA devem ser quebrados em dois locais diferentes, seguidos pela reunião das extremidades quebradas, para produzir um novo rearranjo cromossômico (Figura 17.19, *à esquerda*). Os rearranjos cromossômicos por quebra podem ser induzidos artificialmente por meio da utilização de radiação ionizante. Esse tipo de radiação, notavelmente raios X e gama, é altamente energético e causa diversas quebras bifilamentares no DNA.

Para compreender como os rearranjos cromossômicos são produzidos por meio de quebra, deve-se ter em mente diversos pontos:

1. Cada cromossomo é uma única molécula de DNA bifilamentar.

2. O primeiro evento na produção de um rearranjo cromossômico é a geração de duas ou mais quebras bifilamentares nos cromossomos de uma célula (ver Figura 17.19, *linha superior à esquerda*).
3. As quebras bifilamentares são possivelmente letais, exceto se forem reparadas.
4. Os sistemas de reparo na célula corrigem as quebras bifilamentares por meio da reunião das extremidades quebradas (ver Capítulo 16 para uma discussão detalhada sobre o reparo do DNA).
5. Se as duas extremidades da mesma quebra forem reunidas, a ordem do DNA original é restaurada. Se as extremidades de duas quebras diferentes forem unidas, entretanto, o resultado é um ou outro tipo de rearranjo cromossômico.
6. Os únicos rearranjos cromossômicos que sobrevivem à meiose são aqueles que produzem moléculas de DNA que apresentam um centrômero e dois telômeros. Se um rearranjo produzir um cromossomo com ausência de um centrômero, tal cromossomo **acêntrico** não será atraído para nenhum polo na anáfase da mitose ou meiose e não será incorporado ao núcleo de qualquer progênie. Portanto, os cromossomos acêntricos não são herdados. Se um rearranjo produzir um cromossomo com dois centrômeros (**dicêntrico**), com frequência ele será atraído simultaneamente para os polos opostos na anáfase, formando uma **ponte anafásica**. Os cromossomos com ponte anafásica tipicamente não serão incorporados na célula de qualquer progênie. Se uma quebra cromossômica produzir um cromossomo com ausência de telômero, tal cromossomo não poderá replicar-se adequadamente. Relembre do Capítulo 7 que os telômeros são necessários para iniciar a replicação adequada do DNA nas extremidades (ver Figura 7.26).

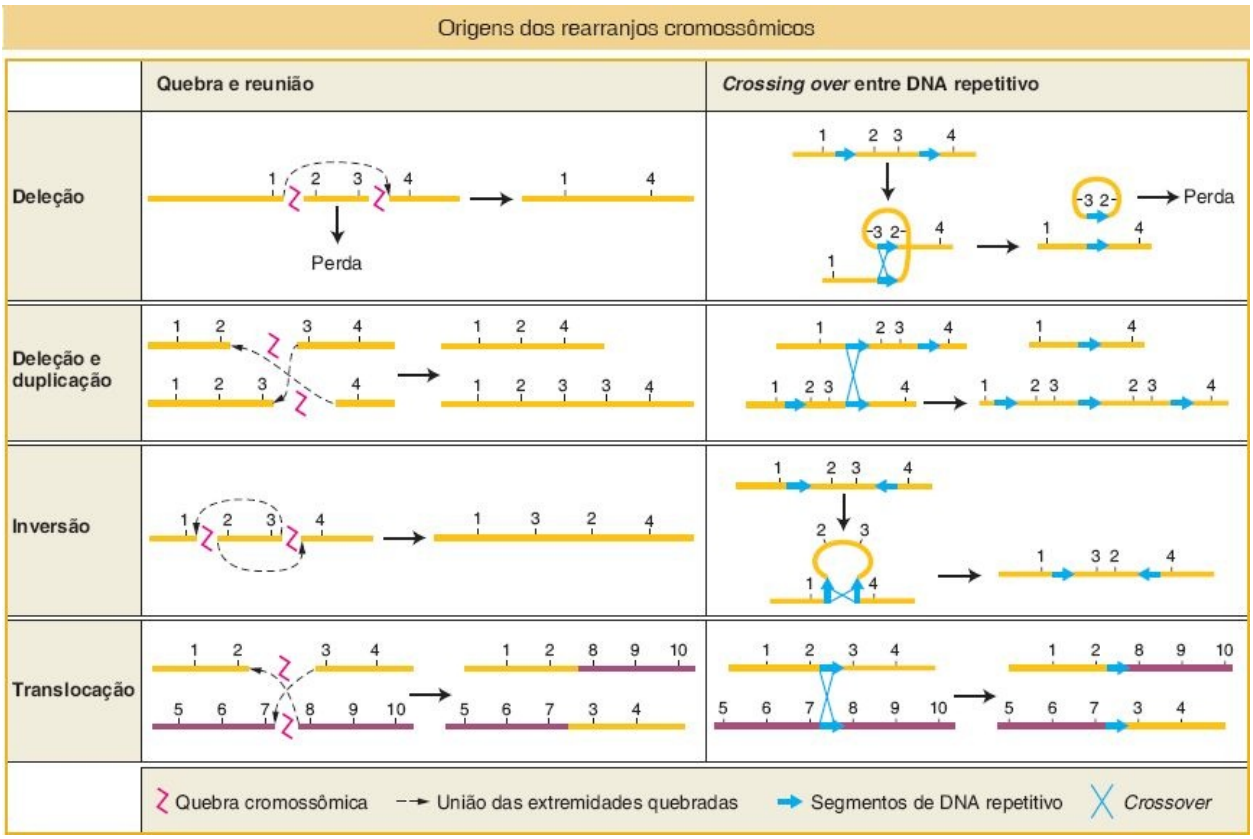


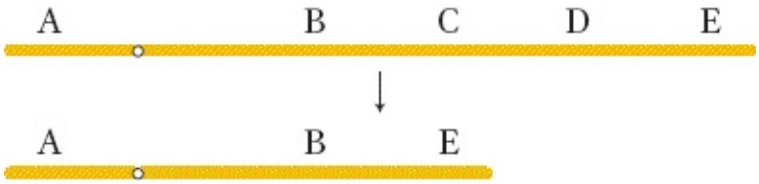
FIGURA 17.19 Cada um dos quatro tipos de rearranjos cromossômicos pode ser produzido por meio de dois mecanismos básicos: quebra cromossômica e reunião ou *crossing over* entre o DNA repetitivo. As regiões cromossômicas são numeradas de 1 a 10. Os cromossomos homólogos são da mesma cor.

7. Se um rearranjo duplicar ou deletar um segmento de um cromossomo, o balanço gênico poderá ser afetado. Quanto maior o segmento que é perdido ou duplicado, maior a probabilidade de que o desbalanceamento gênico cause anormalidades fenotípicas.

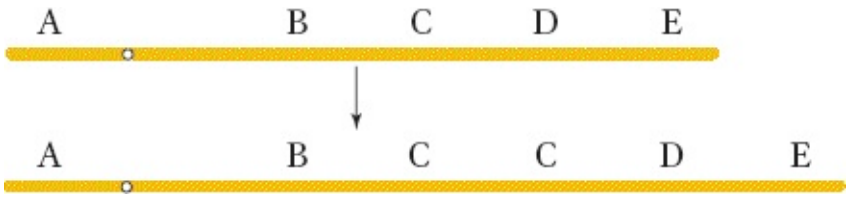
Outra causa importante de rearranjos é o *crossing over* entre segmentos de DNA repetitivo (duplicado). Esse tipo de *crossing over* é denominado **recombinação homóloga não alélica (NAHR)**. Em organismos com sequências repetidas de DNA em um cromossomo ou em cromossomos diferentes, existe uma ambiguidade a respeito de qual das repetições pareará com outra na meiose. Se as sequências pareadas não estiverem nas mesmas posições relativas nos homólogos, o *crossing over* pode produzir cromossomos aberrantes. Deleções, duplicações, inversões e translocações podem, todas, ser produzidas por *crossing over* (ver Figura 17.19,

à direita).

Existem dois tipos gerais de rearranjos: desbalanceado e balanceado. Os **rearranjos desbalanceados** alteram a dosagem gênica de um segmento cromossômico. Assim como na aneuploidia em relação a cromossomos inteiros, a perda de uma cópia de um segmento, ou a adição de uma cópia extra, pode romper o balanço gênico normal. As duas classes simples de rearranjos desbalanceados são as deleções e as duplicações. Uma *deleção* é a perda de um segmento de um braço cromossômico e a justaposição dos dois segmentos em cada lado do segmento deletado, como nesse exemplo, que demonstra a perda do segmento C-D:

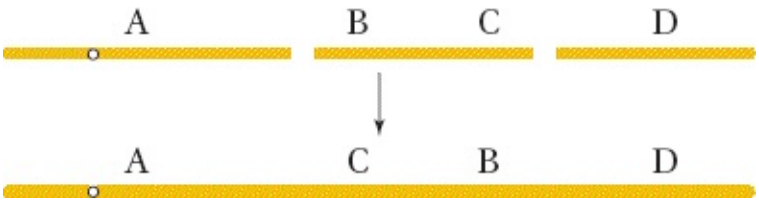


Uma *duplicação* é a repetição de um segmento de um braço cromossômico. No tipo mais simples de duplicação, os dois segmentos estão adjacentes uns aos outros (uma duplicação em *tandem*), assim como nessa duplicação do segmento C:

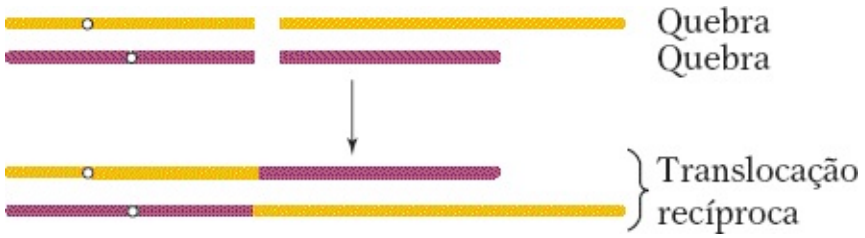


Entretanto, o segmento duplicado pode terminar em uma posição diferente no mesmo cromossomo, ou até mesmo em um cromossomo diferente.

Os **rearranjos balanceados** alteram a ordem dos genes no cromossomo, mas não removem ou duplicam qualquer DNA. As duas classes simples de rearranjos balanceados são as inversões e as translocações recíprocas. Uma *inversão* é um rearranjo no qual um segmento interno de um cromossomo foi quebrado duas vezes, girou 180° e foi reunido.



Uma *translocação recíproca* é um rearranjo no qual dois cromossomos não homólogos são, cada um, quebrados uma vez, criando fragmentos acêntricos, que em seguida podem trocar de lugar:



Por vezes as quebras do DNA que precedem a formação de um rearranjo ocorrem *dentro* dos genes. Quando ocorrem, elas modificam a função do gene, porque parte dele move-se para um novo local e porque nenhum transcrito completo pode ser produzido. Além disso, as sequências de DNA de cada lado das extremidades reunidas de um cromossomo rearranjado são sequências que normalmente não estão justapostas. Por vezes, a junção ocorre de tal modo que a fusão produz um gene híbrido não funcional, composto por partes de dois outros genes.

As seções a seguir consideram as propriedades desses rearranjos balanceados e desbalanceados.

Deleções

Uma deleção é simplesmente a perda de parte de um braço cromossômico. O processo de deleção requer duas quebras cromossômicas para cortar o segmento interveniente. O fragmento deletado não apresenta um centrômero; consequentemente, ele não pode ser puxado para um polo do fuso na divisão celular e é perdido. Os efeitos das deleções dependem de seu tamanho. Uma deleção pequena *dentro* de um gene, denominada **deleção intragênica**, inativa o gene e apresenta o mesmo efeito de outras mutações nulas daquele gene. Se o fenótipo nulo homozigoto for viável (como, por exemplo, no albinismo humano), a

deleção homozigota também será viável. As deleções intragênicas podem ser distinguidas das mutações causadas por alterações em um único nucleotídio, tendo em vista que os genes com as referidas deleções nunca revertem para o tipo selvagem.

Na maior parte desta seção, estaremos lidando com **deleções multigênicas**, nas quais diversos a muitos genes são perdidos. As consequências dessas deleções são mais graves do que aquelas das deleções intragênicas. Se tal deleção ocorrer em homozigose por endogamia (ou seja, se ambos os homólogos apresentarem a mesma deleção), a combinação será sempre letal. Esse fato sugere que todas as regiões dos cromossomos são essenciais para a viabilidade normal e que a eliminação completa de qualquer segmento do genoma é deletéria. Até mesmo um organismo heterozigoto para uma deleção multigênica — ou seja, que apresenta um homólogo normal e um com a deleção — não sobrevive. Principalmente, esse desfecho letal ocorre em virtude do rompimento do balanço gênico normal. Alternativamente, a deleção pode "revelar" alelos recessivos deletérios, possibilitando que as cópias únicas sejam expressas.

> CONCEITO-CHAVE A letalidade das grandes deleções heterozigotas pode ser explicada pelo desbalanceamento gênico e pela expressão de recessivos deletérios.

Deleções pequenas por vezes são viáveis em combinação com um homólogo normal. Tais deleções podem ser identificadas por meio do exame dos cromossomos meióticos ao microscópio. A falha do segmento correspondente no homólogo normal em parear cria uma **alça de deleção** visível (Figura 17.20 A). Em *Drosophila*, as alças de deleção também estão visíveis nos **cromossomos politênicos**. Esses cromossomos são observados nas células de glândulas salivares e em outros tecidos específicos de determinados insetos. Nessas células, os homólogos pareiam e replicam muitas vezes e, assim, cada cromossomo é representado por um espesso feixe de réplicas. Esses cromossomos politênicos são facilmente visíveis e cada um apresenta um

conjunto de bandas escuras de posição e número fixos. Essas bandas atuam como marcos cromossômicos úteis. Um exemplo de um cromossomo politênico no qual um homólogo original carreava uma deleção está demonstrado na Figura 17.20 B. Uma deleção pode ser atribuída a um local específico no cromossomo por meio do exame microscópico dos cromossomos politênicos e da determinação da posição da alça de deleção.

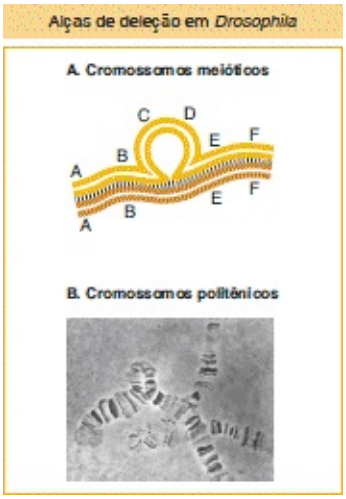
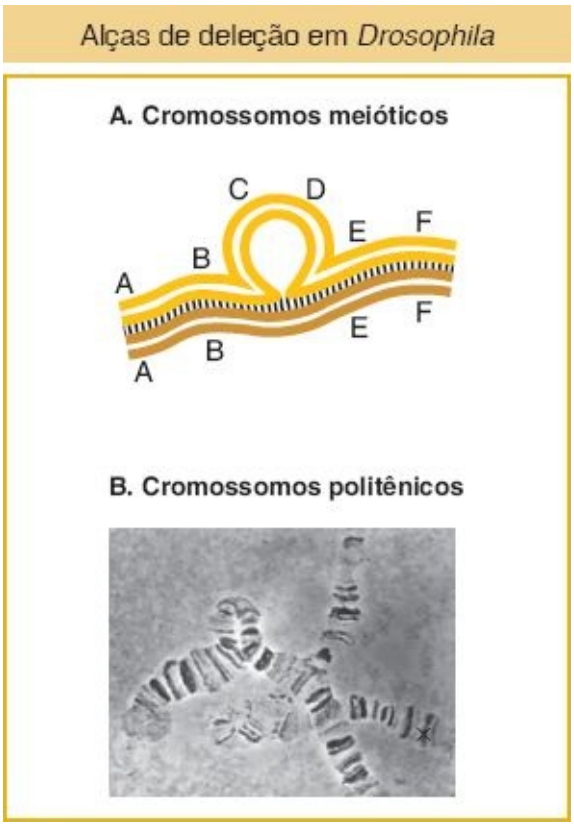


FIGURA 17.20 Na meiose, os cromossomos de uma deleção heterozigota formam uma configuração em alça. A. No pareamento meiótico, o homólogo normal forma uma alça. Os genes nesta alça não apresentam alelos com os quais realizar sinapse. B. Tendo em vista que os cromossomos politênicos de *Drosophila* (observados em glândulas salivares e em outros locais específicos) apresentam padrões de bandeamento específicos, podemos inferir quais bandas estão ausentes no homólogo com a deleção ao observar quais bandas aparecem na alça do homólogo normal. (*B.* William M. Gelbart, Harvard University.)

Outra indicação da presença de uma deleção é que a deleção de um segmento em um homólogo por vezes revela os alelos recessivos presentes no outro homólogo, levando à sua expressão inesperada. Considere, por exemplo, a deleção demonstrada no diagrama a seguir:



Se não houver deleção, espera-se que nenhum dos sete alelos recessivos seja expresso; entretanto, se *b* e *c* forem expressos, então provavelmente ocorreu uma deleção que abrange os genes b^+ e c^+ no outro homólogo. Tendo em vista que os alelos recessivos aparentam estar demonstrando dominância nos referidos casos, o efeito é denominado **pseudodominância**.

No caso reverso — se já conhecemos o local da deleção —, podemos aplicar o efeito da pseudodominância no sentido oposto para mapear as posições dos alelos mutantes. Esse procedimento, denominado **mapeamento de deleção**, pareia as mutações em face de um conjunto de deleções sobrepostas definidas. Um exemplo em *Drosophila* está demonstrado na Figura 17.21. Nesse diagrama, o mapa de recombinação está demonstrado na parte superior, marcado com as distâncias em unidades de mapa a partir da extremidade esquerda. As barras vermelhas horizontais abaixo do cromossomo demonstram a extensão das deleções listadas à esquerda. Cada deleção está pareada com cada mutação em teste e o fenótipo é observado para verificar se a mutação é pseudodominante. A mutação *pn* (*prune*), por exemplo, demonstra pseudodominância apenas com a

deleção 264-38 e esse resultado determina sua localização na região 2D-4 a 3A-2. Entretanto, *fa* (*facet*) demonstra pseudominância com todas as deleções, com exceção de duas (258-11 e 258-14); assim, a sua posição pode ser apontada para a banda 3C-7, que é a região que todas as deleções apresentam em comum, com exceção de duas.

> CONCEITO-CHAVE As deleções podem ser reconhecidas por meio das alças de deleção e da pseudodominância.

Os médicos encontram regularmente deleções nos cromossomos humanos. As deleções normalmente são pequenas, mas apresentam efeitos adversos, mesmo quando heterozigotas. As deleções de regiões cromossômicas humanas específicas causam síndromes únicas de anormalidades fenotípicas. Um exemplo é a síndrome *cri du chat*, causada por uma deleção heterozigota da extremidade do braço curto do cromossomo 5 (Figura 17.22). As bandas específicas deletadas na síndrome *cri du chat* são 5p15.2 e 5p15.3, as duas bandas mais distais identificáveis em 5p. (Os braços curto e longo dos cromossomos humanos são tradicionalmente denominados *p* e *q*, respectivamente.) O fenótipo mais característico na síndrome é aquele que dá origem à sua denominação, o choro semelhante ao miado de um gato produzido pelas crianças afetadas. Outras manifestações da síndrome são microencefalia (cabeça anormalmente pequena) e face redonda (fácies de lua cheia). Assim como as síndromes causadas por outras deleções, a síndrome *cri du chat* inclui o retardo mental. As taxas de fatalidade são baixas e muitas pessoas com essa deleção alcançam a fase adulta.

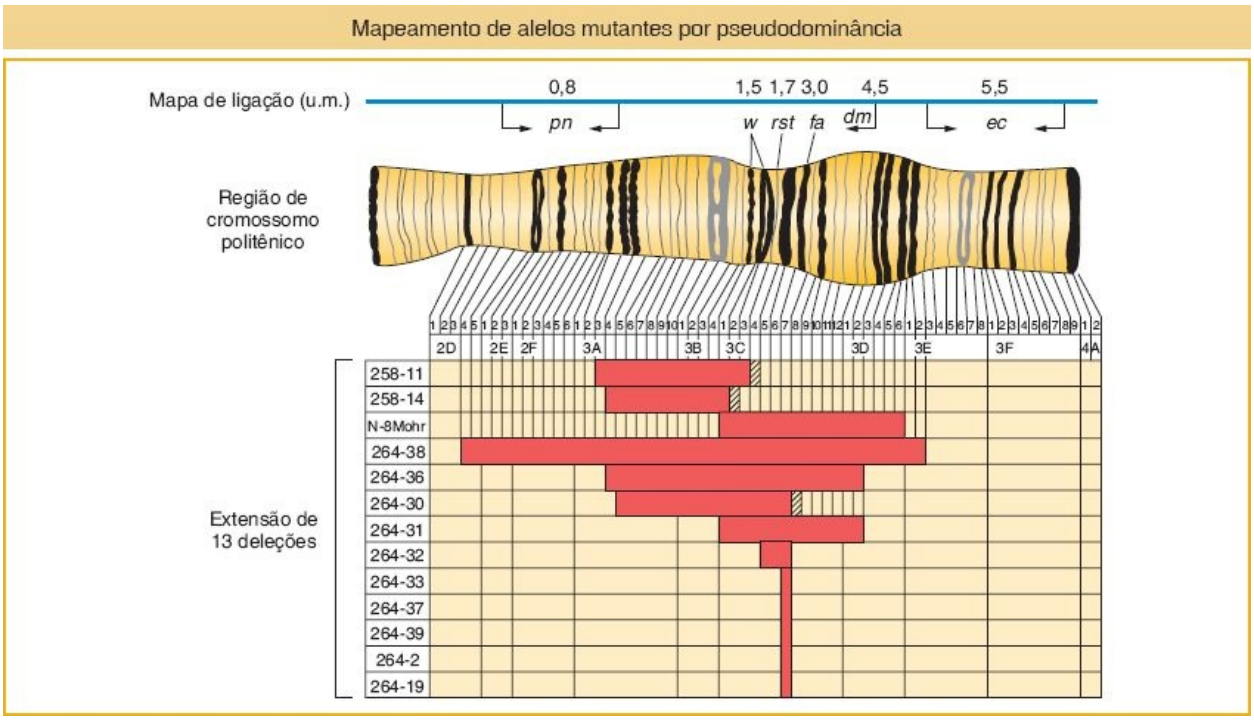


FIGURA 17.21 Uma linhagem heterozigota de *Drosophila* para cromossomos com deleção e normais pode ser utilizada para mapear alelos mutantes. As *barras vermelhas* demonstram a extensão dos segmentos deletados em 13 deleções. Todos os alelos recessivos na mesma região deletada em um cromossomo homólogo serão expressos.

Outro exemplo instrutivo é a síndrome de Williams. Essa síndrome é autossômica dominante e é caracterizada pelo desenvolvimento incomum do sistema nervoso e de determinadas características externas. A síndrome de Williams é observada a uma frequência de aproximadamente 1 em 10.000 pessoas. Os pacientes com frequência apresentam habilidades musicais ou de canto pronunciadas. A síndrome quase sempre é causada por uma deleção de 1,5 Mb em um homólogo do cromossomo 7. A análise da sequência demonstrou que esse segmento contém 17 genes de função conhecida e desconhecida. O fenótipo anormal, portanto, é causado por haploinsuficiência de um ou mais desses 17 genes. A análise da sequência também revela a origem dessa deleção, tendo em vista que a sequência normal é delimitada por cópias repetidas de um gene denominado *PMS*, que codifica uma proteína de reparo do DNA. Conforme vimos, as sequências repetidas podem atuar como substratos para *crossing over* desigual. Um *crossover* entre cópias flanqueadoras de *PMS* nas extremidades

opostas do segmento dos 17 genes leva a uma duplicação (não observada) e uma deleção na síndrome de Williams, conforme demonstrado na Figura 17.23.

A maior parte das deleções humanas, tais como aquelas que acabamos de considerar, surge espontaneamente nas gônadas de um genitor normal de uma pessoa afetada; portanto, normalmente não são observados sinais de deleções nos cromossomos dos genitores. Menos comumente, os indivíduos que contêm deleções aparecem na descendência de um indivíduo que apresenta um rearranjo balanceado de cromossomos não detectado. Por exemplo, a síndrome *cri du chat* pode resultar de um genitor heterozigoto para uma translocação recíproca, tendo em vista que (conforme veremos) a segregação produz deleções. As deleções também podem resultar de recombinação em um heterozigoto que apresenta uma inversão pericêntrica (uma inversão que abrange o centrômero) em um cromossomo. Ambos os mecanismos serão detalhados posteriormente no capítulo.

Animais e plantas demonstram diferenças na sobrevivência de gametas ou de descendência que contêm deleções. Um animal macho com uma deleção em um cromossomo produz espermatozoides que carreiam um ou outro dos dois cromossomos em números aproximadamente iguais. Esses espermatozoides aparentemente funcionam, até uma determinada medida, independentemente de seu conteúdo genético. Em plantas diploides, por outro lado, o pólen produzido por uma deleção heterozigota é de dois tipos: pólen funcional que carreia o cromossomo normal e pólen não funcional (abortado) que carreia o homólogo deficiente. Portanto, as células do pólen aparentam ser sensíveis às alterações na quantidade de material cromossômico e essa sensibilidade pode atuar para extirpar as deleções. Esse efeito é análogo à sensibilidade do pólen à aneuploidia de um cromossomo inteiro, descrita anteriormente neste capítulo. Contrariamente às células de espermatozoides de animais, cuja atividade metabólica depende de enzimas que já foram depositadas nelas durante a sua formação, as células de pólen devem germinar e em seguida produzir um longo tubo de pólen, que cresce para fertilizar o óvulo. Esse crescimento requer que a célula de pólen produza grandes quantidades de proteínas, tornando-a, assim, sensível às anormalidades genéticas em seu próprio núcleo. Os óvulos das plantas, ao contrário, são razoavelmente tolerantes às deleções, presumivelmente em virtude de receberem

sua nutrição dos tecidos maternos circundantes.

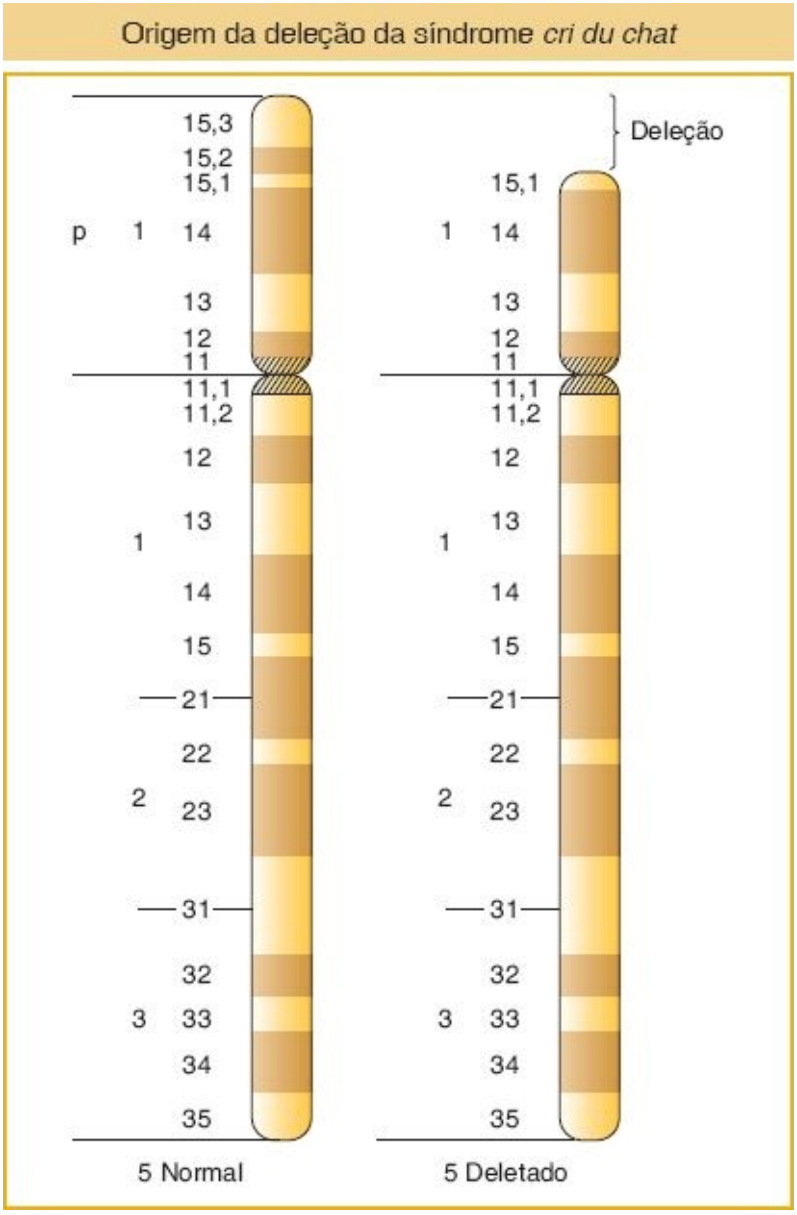


FIGURA 17.22 A síndrome *cri du chat* é causada pela perda da extremidade do braço curto de um dos homólogos do cromossomo 5.

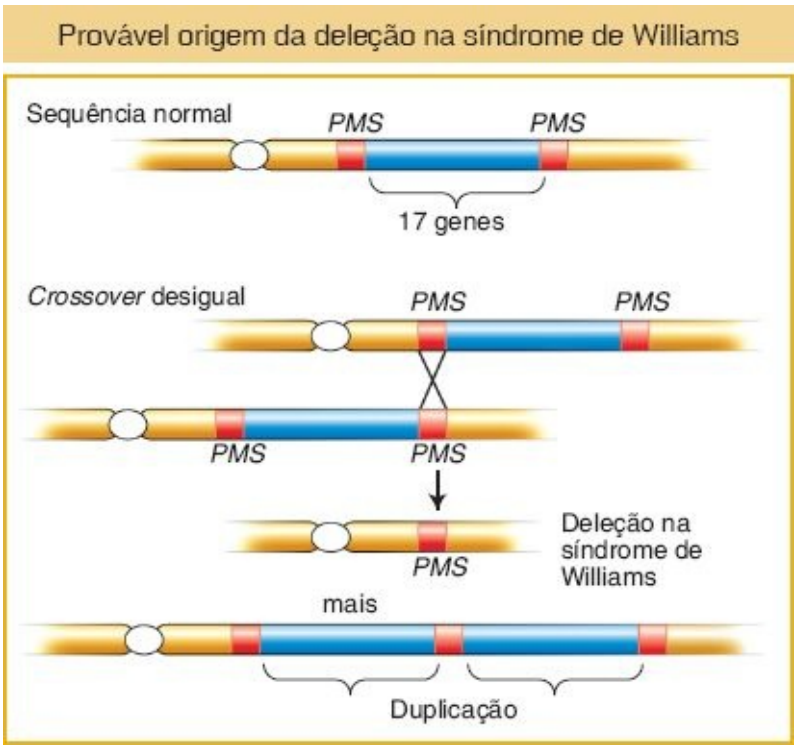


FIGURA 17.23 Um *crossover* entre os genes flanqueadores repetitivos esquerdo e direito resulta em dois rearranjos recíprocos, um dos quais corresponde à deleção na síndrome de Williams.

Duplicações

Os processos de mutação cromossômica por vezes produzem uma cópia extra de alguma região cromossômica. As regiões duplicadas podem estar localizadas adjacentes entre si — denominadas **duplicação em *tandem*** — ou a cópia extra pode estar localizada em algum outro local no genoma — denominada **duplicação insercional**. Uma célula diploide que contém uma duplicação apresentará três cópias da região cromossômica em questão: duas em um conjunto cromossômico e uma no outro — um exemplo de uma duplicação heterozigota. Na prófase meiótica, os heterozigotos com duplicação em *tandem* demonstram uma alça formada pela região extra não pareada.

As duplicações sintéticas de cobertura conhecida podem ser utilizadas para o mapeamento gênico. Em haploides, por exemplo, uma linhagem cromossomicamente normal que carreia uma nova mutação recessiva *m* pode ser cruzada com linhagens que contêm numerosos rearranjos geradores de

duplicações (p. ex., translocações e inversões pericêntricas). Em qualquer cruzamento, se alguma progênie com duplicação apresentar o fenótipo recessivo, a duplicação não abrangerá o gene *m*, tendo em vista que, se ela abrangesse, o seu segmento extra mascararia o alelo recessivo *m*.

Análises das sequências de DNA genômico revelaram um alto nível de duplicações em seres humanos e na maior parte dos organismos-modelo. Repetições de sequências simples, que se estendem por todo o genoma e são úteis como marcadores moleculares no mapeamento, foram discutidas nos capítulos anteriores. Entretanto, outra classe de duplicações tem por base unidades duplicadas que são muito maiores do que as repetições de sequências simples. As duplicações nessa classe são denominadas **duplicações segmentares**. As unidades duplicadas nas duplicações segmentares variam de 10 a 50 quilobases de comprimento e abrangem genes inteiros e as regiões entre eles. A extensão das duplicações segmentares está demonstrada na Figura 17.24, na qual a maior parte das duplicações está dispersa, mas existem alguns casos em *tandem*. Outra propriedade demonstrada na Figura 17.24 é que a dispersão das unidades duplicadas está principalmente no mesmo cromossomo, não entre cromossomos. A origem das duplicações segmentares ainda não é conhecida.

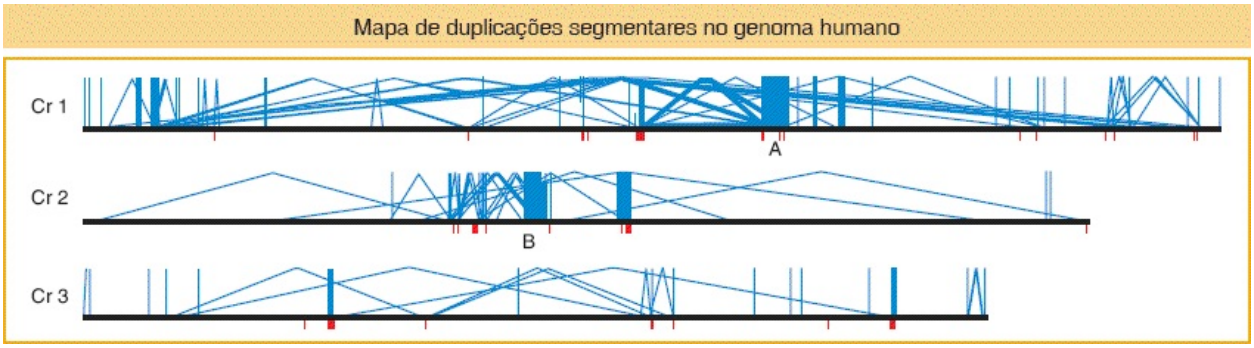


FIGURA 17.24 O mapa dos cromossomos humanos 1, 2 e 3 demonstra as posições das duplicações com tamanho superior a 10 quilobases. As *linhas de conexão azuis* demonstram duplicações intracromossômicas (a maioria). As duplicações intracromossômicas estão demonstradas com *barras vermelhas*. As letras A e B indicam *hotspots* nos quais a recombinação de duplicações deu origem a distúrbios genéticos. (Dados de J. A. Bailey *et al.*, "Recent Segmental Duplications in the Human Genome", *Science* 297, 2002, 1003-1007.)

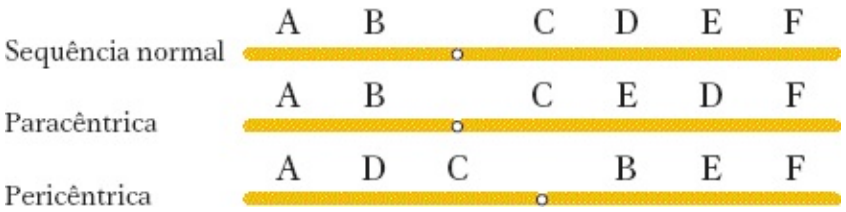
Acredita-se que as duplicações segmentares apresentem um papel importante

como substratos para a recombinação homóloga não alélica, conforme demonstrado na Figura 17.19. O *crossing over* entre duplicações segmentares pode levar a diversos rearranjos cromossômicos. Esses rearranjos aparentam ter sido importantes na evolução, uma vez que algumas inversões importantes, que são diferenças-chave entre as sequências humanas e de primatas, quase certamente têm origem na NAHR (recombinação homóloga não alélica). Também aparenta ser provável que a NAHR seja responsável pelos rearranjos que causam algumas doenças humanas. Os *loci* das referidas doenças estão em *hotspots* de duplicação segmentar; exemplos dos referidos *loci* estão demonstrados na Figura 17.24.

Observamos que, em alguns organismos, como os poliploides, o genoma atual evoluiu como resultado de uma duplicação de todo o genoma ancestral. Após a ocorrência de uma duplicação do genoma inteiro, todos os genes são duplicados. Esses genes duplicados são uma fonte de algumas das duplicações segmentares observadas nos genomas. Um caso bem-estudado é *Saccharomyces cerevisiae*. A evolução desse genoma foi analisada por meio da comparação da sequência de todo o genoma de *S. cerevisiae* com o de outra levedura, *Kluyveromyces*, cujo genoma é semelhante ao do ancestral de levedura. Aparentemente, no período da evolução de *Saccharomyces*, o genoma ancestral semelhante ao de *Kluyveromyces* foi duplicado e assim existiam dois conjuntos, cada um contendo o genoma inteiro. Após a ocorrência da duplicação, muitas cópias gênicas foram perdidas de um conjunto ou do outro, e os conjuntos remanescentes foram rearranjados, resultando no genoma atual de *Saccharomyces*. Esse processo está reconstruído na Figura 17.25.

Inversões

Observamos que, para criar uma inversão, um segmento de um cromossomo é cortado, girado e reinserido. As inversões são de dois tipos básicos. Se o centrômero estiver fora da inversão, diz-se que a inversão é **paracêntrica**. As inversões que abrangem o centrômero são **pericêntricas**.



Tendo em vista que as inversões são rearranjos balanceados, elas não alteram a quantidade geral de material genético e, assim, não resultam em desbalanceamento gênico. Os indivíduos com inversões em geral são normais, se não houver quebras dentro dos genes. Uma quebra que rompe um gene produz uma mutação que pode ser detectável como um fenótipo anormal. Se o gene apresenta uma função essencial, o ponto de quebra então atua como uma mutação letal ligada à inversão. Nesse caso, a inversão não pode dar origem à homozigose. Entretanto, muitas inversões podem se tornar homozigotas e, além disso, as inversões podem ser detectadas em organismos haploides. Nesses casos, os pontos de quebra da inversão claramente não estão em regiões essenciais. Algumas das possíveis consequências da inversão no nível do DNA estão demonstradas na Figura 17.26.

A maior parte das análises de inversões é realizada em células diploides que contêm um conjunto cromossômico normal mais um conjunto que carreia a inversão. Esse tipo de célula é denominado **heterozigota para inversão**, mas observe que essa designação não implica que qualquer *locus* gênico seja heterozigoto; em vez disso, significa que um conjunto cromossômico normal e um anormal estão presentes. A localização do segmento invertido com frequência pode ser detectada microscopicamente. Na meiose, um cromossomo se torce nas extremidades da inversão, formando uma alça para parear com seu homólogo não torcido; desse modo, os homólogos pareados formam uma **alça de inversão** visível (Figura 17.27).

Em uma inversão *paracêntrica*, o *crossing over* na alça de inversão durante a meiose conecta os centrômeros homólogos em uma **ponte dicêntrica**, enquanto também produz um **fragmento acêntrico** (Figura 17.28). Em seguida, na medida em que os cromossomos se separam na anáfase I, os centrômeros permanecem ligados pela ponte. O fragmento acêntrico não consegue se alinhar ou se

movimentar; consequentemente, ele é perdido. A tensão finalmente quebra a ponte dicêntrica, formando dois cromossomos com deleções terminais. Os gametas que contêm tais cromossomos, ou os zigotos que eles finalmente venham a formar, provavelmente serão inviáveis. Portanto, um evento de *crossover*, que normalmente dá origem à classe recombinante de produtos meióticos, é, em vez disso, letal para aqueles produtos. O resultado geral é uma frequência drasticamente mais baixa de recombinantes viáveis. De fato, em relação aos genes dentro da inversão, a frequência de recombinantes está próxima de zero. (Ela não é exatamente zero em virtude de os raros *crossovers* duplos entre apenas duas cromátides serem viáveis.) Em relação aos genes que flanqueiam a inversão, a frequência de recombinantes é reduzida em proporção ao tamanho da inversão, tendo em vista que, em uma inversão mais longa, existe maior probabilidade de ocorrência de um *crossover* dentro dela e da produção de um produto meiótico inviável.

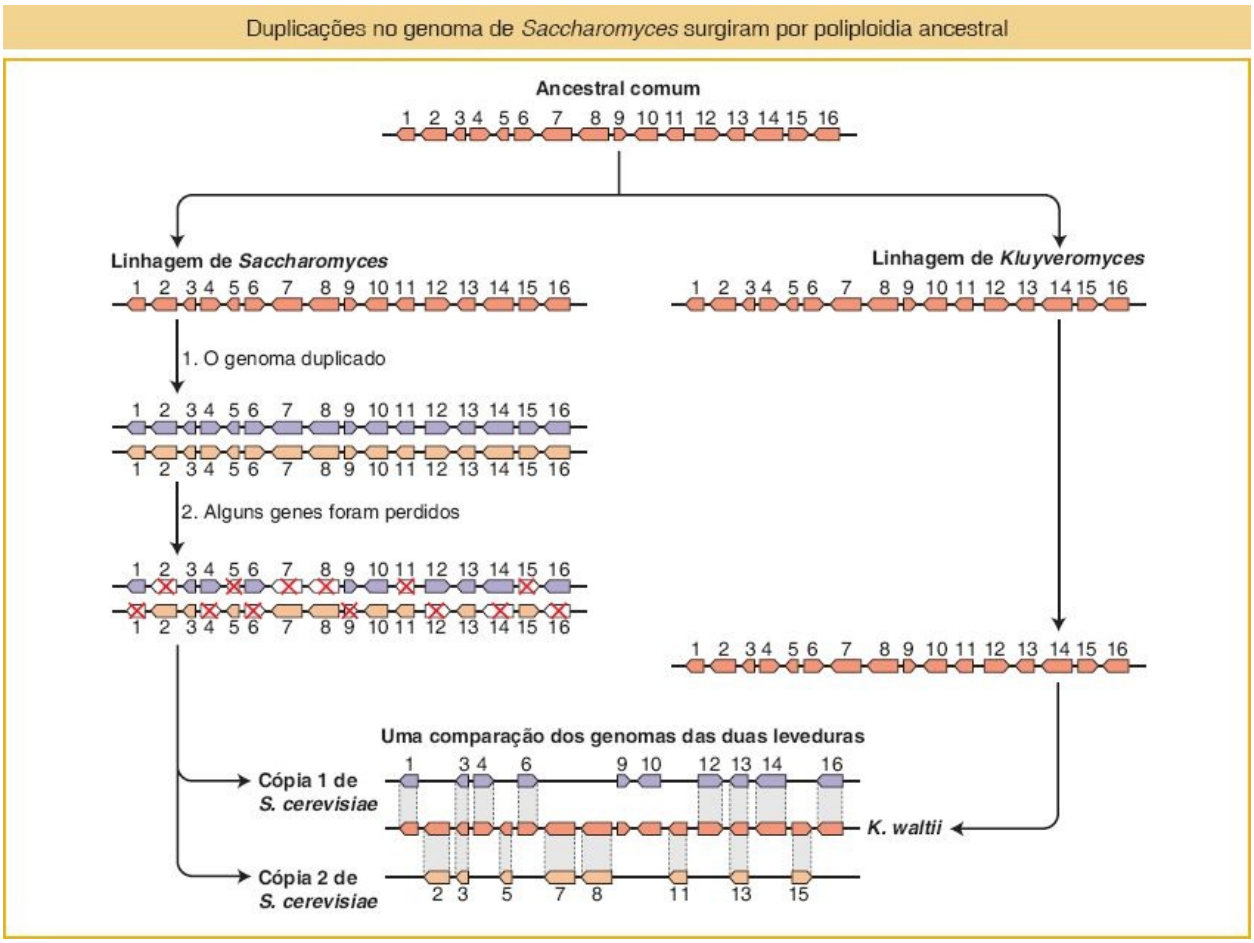


FIGURA 17.25 Um ancestral comum semelhante à levedura *Kluyveromyces* moderna duplicou seu genoma (**1**). Alguns genes foram perdidos (**2**). Os genes duplicados, tais como 3 e 13, estão na mesma ordem relativa. O painel inferior compara os dois genomas modernos. (Dados da Figura 1, Manolis Kellis, Bruce W. Birren e Eric S. Lander, "Proof and Evolutionary Analysis of Ancient Genome Duplication in the Yeast Saccharomyces cerevisiae", *Nature* 428, 8 de abril de 2004, direitos autorais Nature Publishing Group.)

Em uma inversão *pericêntrica* heterozigota, o efeito genético líquido é o mesmo daquele de uma inversão paracêntrica — os produtos do *crossover* não são recuperados —, mas os motivos são diferentes. Em uma inversão pericêntrica, os centrômeros estão contidos dentro da região invertida. Consequentemente, os cromossomos envolvidos no *crossover* se separam do modo normal, sem a criação de uma ponte (Figura 17.29). Entretanto, o *crossover* produz cromátides que contêm uma duplicação e uma deleção em diferentes partes do cromossomo. Nesse caso, se um gameta que carrega um cromossomo com *crossover* é fertilizado, o zigoto morre em virtude do desbalanceamento

gênico. Novamente, o resultado é que apenas as cromátides que não sofreram *crossover* estão presentes na progênie viável. Portanto, o valor da frequência de recombinação de genes em uma inversão pericêntrica também é zero.

As inversões também afetam a recombinação de outro modo. Os heterozigotos com inversão frequentemente apresentam problemas mecânicos de pareamento na região da inversão. A alça de inversão causa uma grande distorção, que pode se estender além da própria alça. Essa distorção reduz a oportunidade de realização de *crossing over* nas regiões vizinhas.

As inversões podem causar uma diversidade de alterações estruturais no DNA

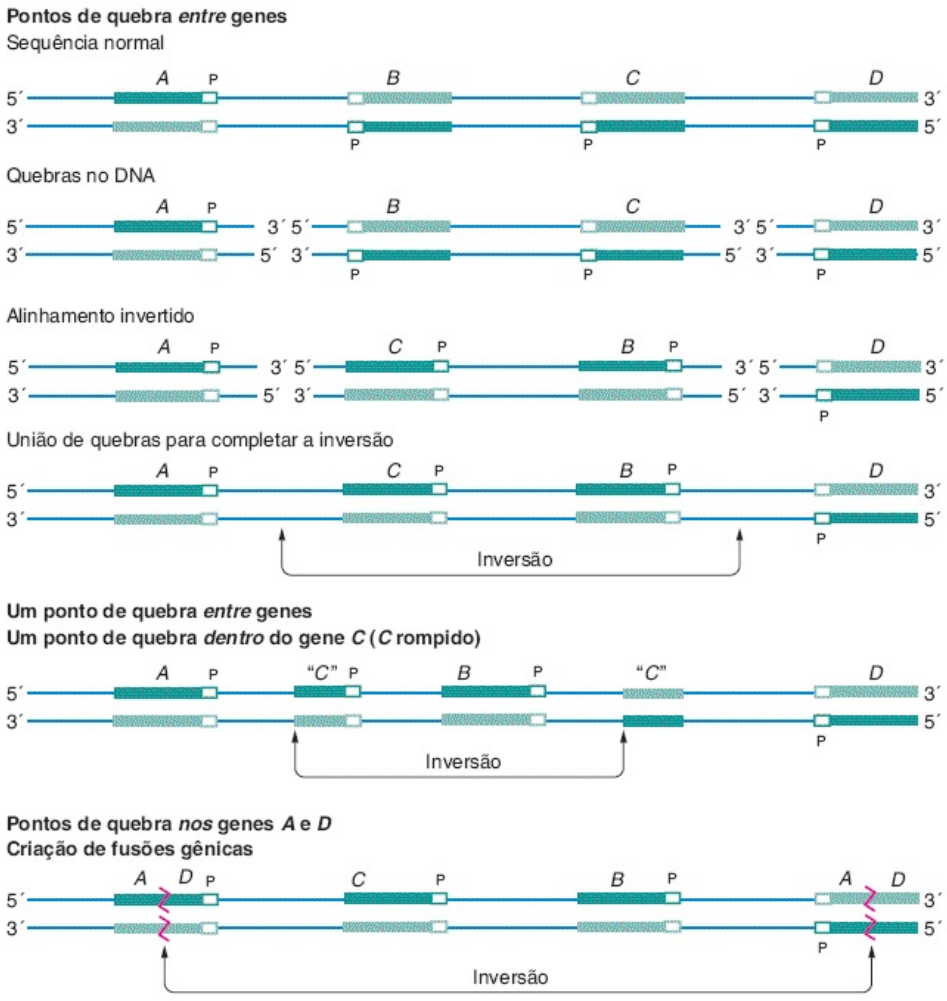


FIGURA 17.26 Uma inversão pode não apresentar nenhum efeito sobre os genes, pode romper um gene ou

pode fundir partes de dois genes, dependendo do local do ponto de quebra. Os genes estão representados por *A*, *B*, *C* e *D*. O filamento-molde é *verde-escuro*; o filamento não molde é *verde-claro*; as *linhas vermelhas* denteadas indicam onde as quebras no DNA produziram fusões gênicas (*A* com *D*) após a inversão e a reunião. A letra P faz referência ao promotor; as *setas* indicam as posições dos pontos de quebra.

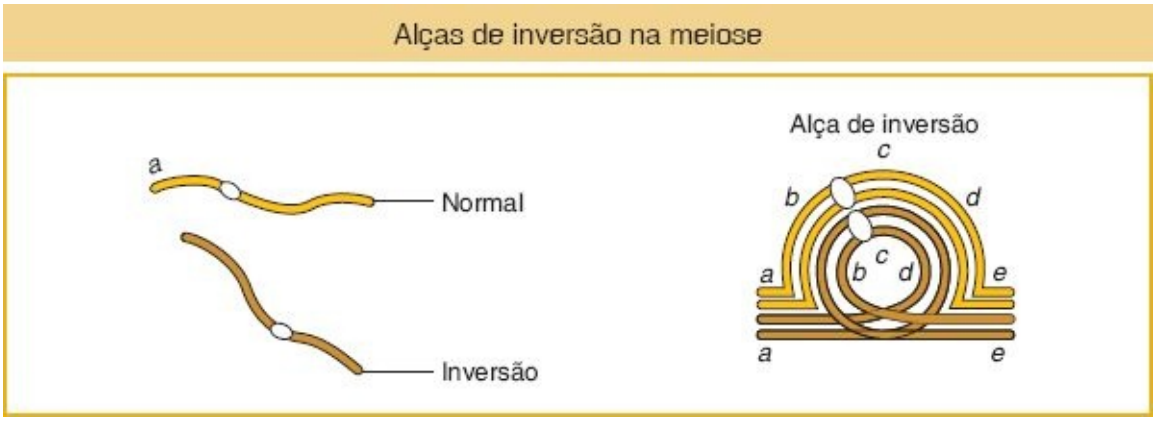


FIGURA 17.27 Os cromossomos de heterozigotos para inversão pareiam em uma alça na meiose.

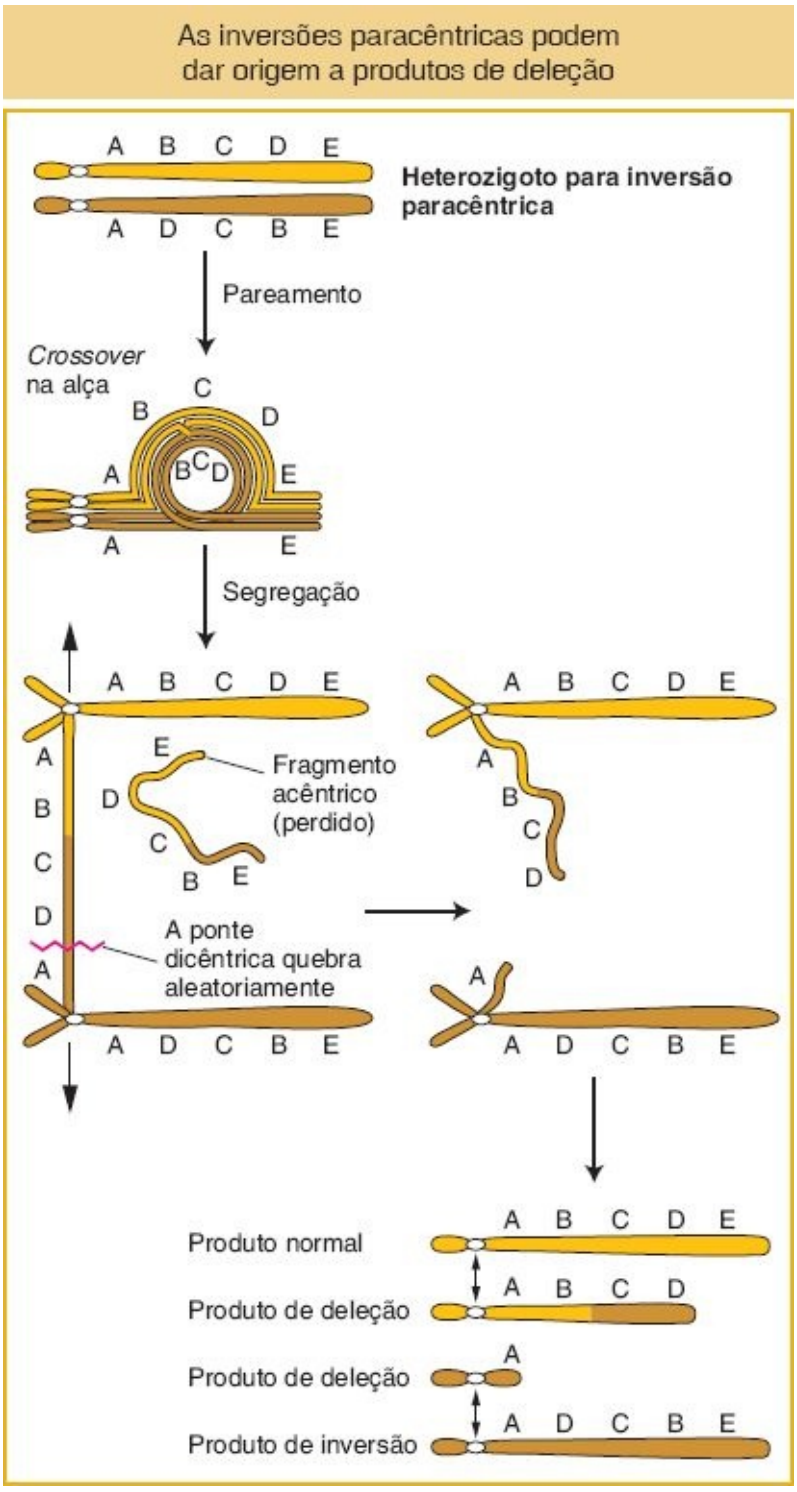


FIGURA 17.28 Um *crossover* na alça de um heterozigoto com inversão paracêntrica dá origem a cromossomos que contêm deleções.

Consideraremos um exemplo dos efeitos de uma inversão sobre a frequência de

recombinantes. Um espécime de *Drosophila* do tipo selvagem de uma população natural é cruzado com um estoque de laboratório homozigoto recessivo *dp cn/dp cn*. (O alelo *dp* codifica asas *dumpy* e *cn* codifica olhos *cinnabar*. Os dois genes sabidamente estão distantes 45 unidades de mapa no cromossomo 2.) A geração F_1 é do tipo selvagem. Quando uma fêmea da F_1 é cruzada com o genitor recessivo, a progênie é:

250	tipo selvagem	*+ +/dp cn*
246	asas *dumpy* e olhos *cinnabar*	*dp cn/dp cn*
5	asas *dumpy*	*dp +/dp cn*
7	olhos *cinnabar*	*+ cn/dp cn*

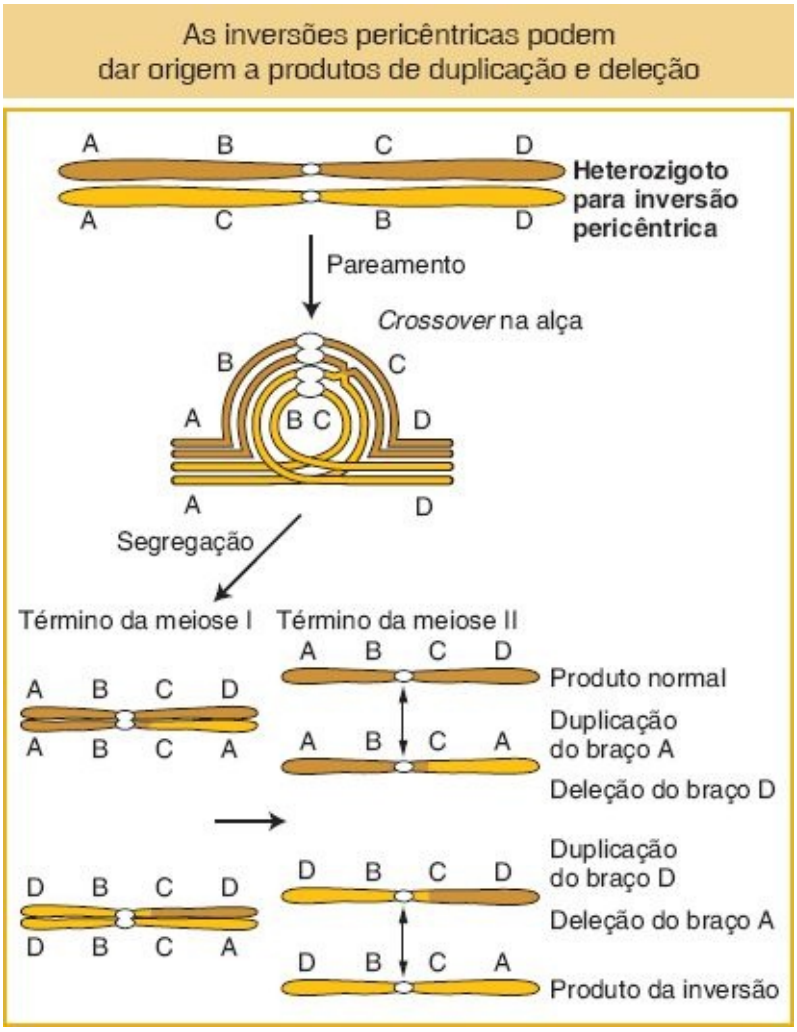


FIGURA 17.29 Um *crossover* na alça de um heterozigoto para inversão pericêntrica dá origem a cromossomos que contêm duplicações e deleções.

Nesse cruzamento, que é efetivamente um cruzamento-teste di-híbrido, espera-se que 45% da progênie seja *dumpy* ou *cinnabar* (elas constituem as classes de *crossover*), mas apenas 12 de 508, aproximadamente 2%, são obtidas. Algo está reduzindo o *crossing over* nessa região e uma provável explicação é uma inversão que abrange a maior parte da região *dp-cn*. Tendo em vista que a FR esperada teve por base as medições realizadas em linhagens de laboratório, a mosca do tipo selvagem da natureza era a fonte mais provável do cromossomo invertido. Portanto, o cromossomo 2 na F_1 pode ser representado como segue:

As inversões pericêntricas também podem ser detectadas microscopicamente por meio das razões do novo braço. Considere a inversão pericêntrica a seguir:

Observe que a razão do comprimento do braço longo e do braço curto foi alterada de aproximadamente 4:1 para aproximadamente 1:1 pela inversão. As inversões paracêntricas não alteram a razão do braço, mas podem ser detectadas microscopicamente por meio da observação das alterações no bandeamento ou de outros marcos cromossômicos, se disponíveis.

> CONCEITO-CHAVE As principais características diagnósticas das inversões heterozigotas são as alças de inversão, a frequência de recombinantes reduzida e a fertilidade reduzida em virtude dos produtos meióticos desbalanceados ou deletados.

Em alguns sistemas-modelo experimentais, notavelmente em *Drosophila* e no nematódeo *Caenorhabditis elegans*, as inversões são utilizadas como balanceadoras. Um cromossomo **balanceador** contém *múltiplas* inversões; assim, quando ele é combinado com o cromossomo do tipo selvagem correspondente, pode não haver produtos de *crossover* viáveis. Em algumas análises, é importante manter um estoque com todos os alelos em um cromossomo juntos. O geneticista cria indivíduos portadores de genomas que combinam um referido cromossomo com um balanceador. Essa combinação elimina *crossovers* e, assim, apenas as combinações parentais aparecem na progênie. Por conveniência, os cromossomos balanceadores são marcados com uma mutação morfológica dominante. O

marcador possibilita que o geneticista rastreie a segregação de todo balanceador ou de seu homólogo normal ao observar a presença ou a ausência do marcador.

Translocações recíprocas

Existem diversos tipos de translocações, mas aqui consideramos apenas as translocações recíprocas, o tipo mais simples. Relembre que, para formar uma translocação recíproca, dois cromossomos realizam permuta de fragmentos acêntricos criados por duas quebras cromossômicas simultâneas. Assim como com outros rearranjos, a meiose em heterozigotos que apresentam dois cromossomos translocados e seus correspondentes normais produz configurações características. A Figura 17.30 ilustra a meiose em um indivíduo que é heterozigoto para uma translocação recíproca. Observe a configuração do pareamento em forma de cruz. Tendo em vista que a lei da distribuição independente está atuando, existem dois padrões comuns de segregação. Utilizaremos N_1 e N_2 para representar os cromossomos normais e T_1 e T_2, os cromossomos translocados. A segregação de cada um dos cromossomos estruturalmente normais com um dos translocados (T_1 + N_2 e T_2 + N_1) é denominada **segregação adjacente 1**. Cada um dos dois produtos meióticos é deficiente para um braço diferente da cruz e apresenta uma duplicata do outro. Esses produtos são inviáveis. Por outro lado, os dois cromossomos normais podem segregar juntos, assim como farão as partes recíprocas dos translocados, para gerar os produtos N_1 + N_2 e T_1 + T_2. Esse padrão de segregação é denominado **segregação alternada**. Esses produtos são, ambos, balanceados e viáveis.

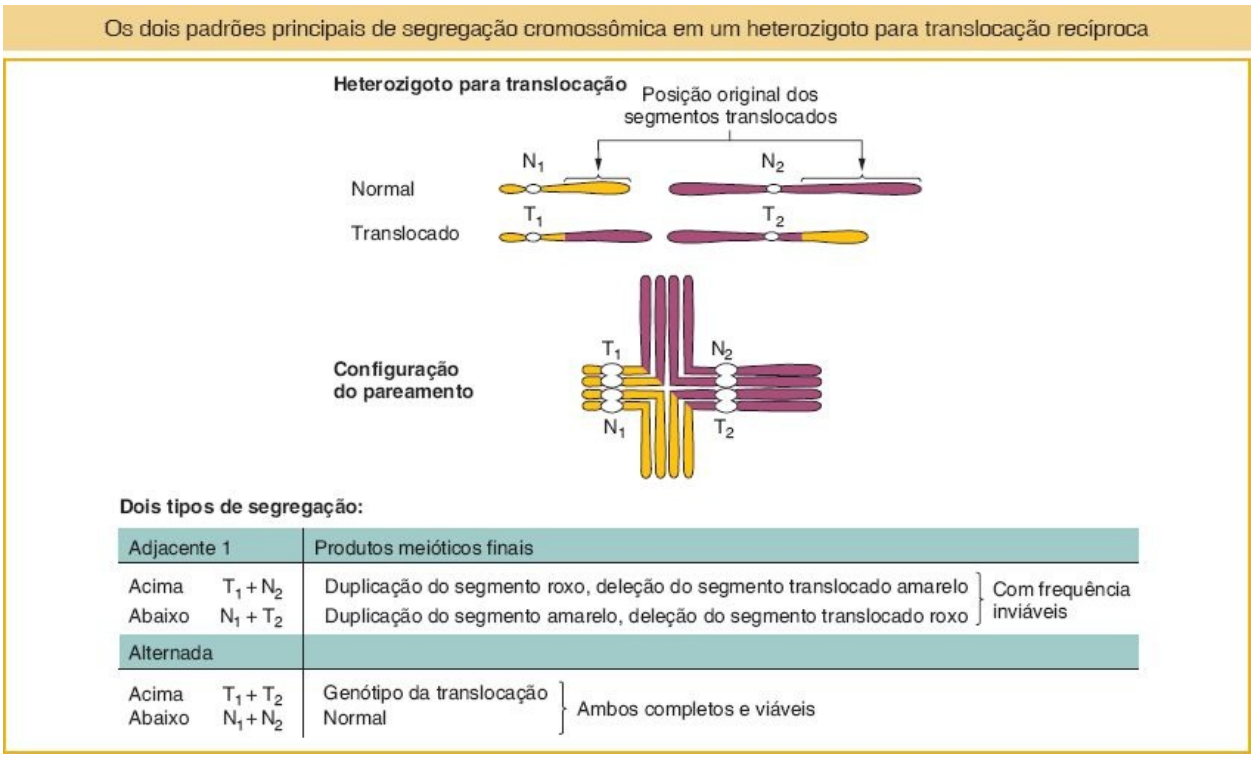


FIGURA 17.30 Os cromossomos em segregação de um heterozigoto para translocação recíproca formam uma configuração de pareamento em cruz. Os dois padrões de segregação mais comumente encontrados que resultam são o "adjacente 1", com frequência inviável, e o "alternado", viável. N_1 e N_2, cromossomos não homólogos normais; T_1 e T_2, cromossomos translocados. "Acima" e "Abaixo" designam os polos opostos para os quais os homólogos migram na anáfase I.

As segregações adjacente 1 e alternada são iguais em número e, assim, metade da população geral de gametas será não funcional, uma condição conhecida como **semiesterilidade**. A semiesterilidade é uma ferramenta diagnóstica importante para a identificação de heterozigotos com translocação. Entretanto, a semiesterilidade é definida de modo diferente para plantas e animais. Em plantas, os 50% dos produtos meióticos provenientes da segregação adjacente 1 em geral abortam no estágio gamético (Figura 17.31). Em animais, esses produtos são viáveis como gametas, mas letais para os zigotos que eles produzem na fertilização.

Relembre que os heterozigotos para inversões também podem demonstrar alguma redução na fertilidade, mas em uma quantidade dependente do tamanho da região afetada. A redução exata de 50% nos gametas ou zigotos viáveis

normalmente é uma indicação diagnóstica confiável de translocação.

Geneticamente, os genes em cromossomos translocados atuam como se estivessem ligados se seus *loci* estiverem próximos do ponto de quebra da translocação. A Figura 17.32 demonstra uma translocação heterozigota que foi estabelecida por meio do cruzamento de um indivíduo *a/a*; *b/b* com um homozigoto para translocação que contém os alelos do tipo selvagem. Quando o heterozigoto é submetido ao cruzamento-teste, são criados recombinantes, mas eles não sobrevivem, tendo em vista que carreiam genomas desbalanceados (duplicações e deleções). A única progênie viável é aquela que contém os genótipos parentais; assim, é observada a ligação entre *loci* que originalmente estavam em cromossomos diferentes. A aparente ligação dos genes normalmente conhecidos por estarem em cromossomos não homólogos separados — por vezes denominada **pseudoligação** — é uma indicação diagnóstica genética da presença de uma translocação.

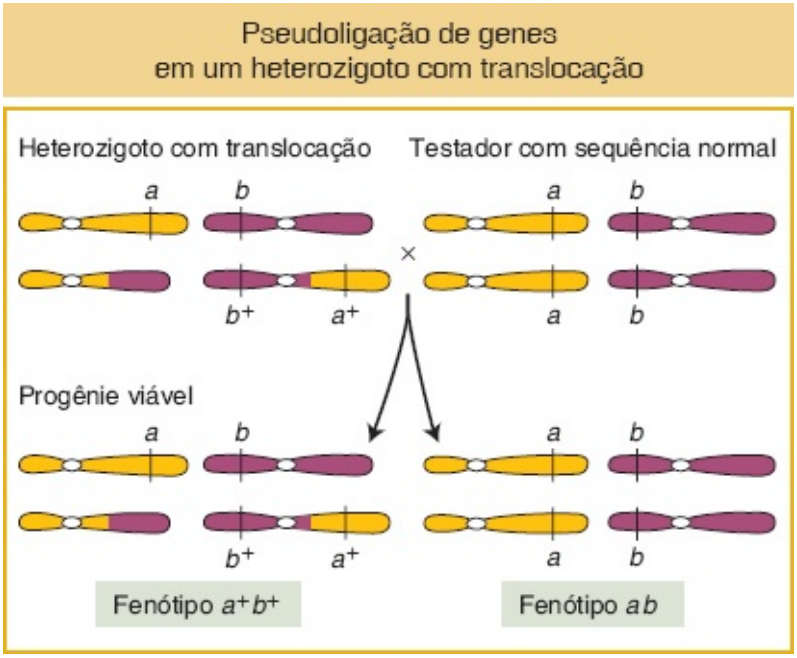


FIGURA 17.31 Pólen de uma planta do milho semiestéril. Os grãos de pólen claros contêm produtos meióticos cromossomicamente desbalanceados de um heterozigoto para translocação recíproca. Os grãos de pólen opacos, que contêm cromossomos com genótipo da translocação completa ou normais, são funcionais na fertilização e no desenvolvimento. (William Sheridan.)

FIGURA 17.32 Quando um fragmento translocado carrega um gene marcador, esse marcador pode demonstrar ligação com genes no outro cromossomo.

> CONCEITO-CHAVE As translocações recíprocas heterozigotas são diagnosticadas geneticamente por meio da semiesterilidade e pela aparente ligação de genes cujos *loci* normais estão em cromossomos separados.

Translocações robertsonianas

Retornaremos para a família da criança com síndrome de Down, introduzida no início do capítulo. O nascimento pode, na verdade, ter sido uma coincidência — afinal, coincidências acontecem. Entretanto, o aborto fornece uma indicação de que pode estar acontecendo algo mais. Uma grande proporção de abortos espontâneos carreia anormalidades cromossômicas, de tal modo que talvez seja esse o caso nesse exemplo. Nesse caso, o casal pode ter tido duas concepções com mutações cromossômicas, o que seria muito improvável, exceto se houvesse uma causa comum. Entretanto, uma pequena proporção de casos de síndrome de Down sabidamente resulta de uma translocação em um dos genitores. Vimos que as translocações podem produzir progênie que apresente material extra de uma parte do genoma e, assim, uma translocação que envolva o cromossomo 21 pode produzir progênie que apresente material extra daquele cromossomo. Na

síndrome de Down, a translocação responsável é de um tipo denominado *translocação robertsoniana*. Ela produz progênie que carrega uma cópia extra quase completa do cromossomo 21. A translocação e sua segregação estão ilustradas na Figura 17.33. Observe que, além dos complementos que causam a síndrome de Down, são produzidos outros complementos cromossômicos aberrantes, cuja maior parte é abortada. Em nosso exemplo, o homem pode apresentar essa translocação, que ele pode ter herdado de sua avó. Para confirmar essa possibilidade, seus cromossomos são verificados. Seu filho não afetado pode apresentar cromossomos normais, ou pode ter herdado essa translocação.

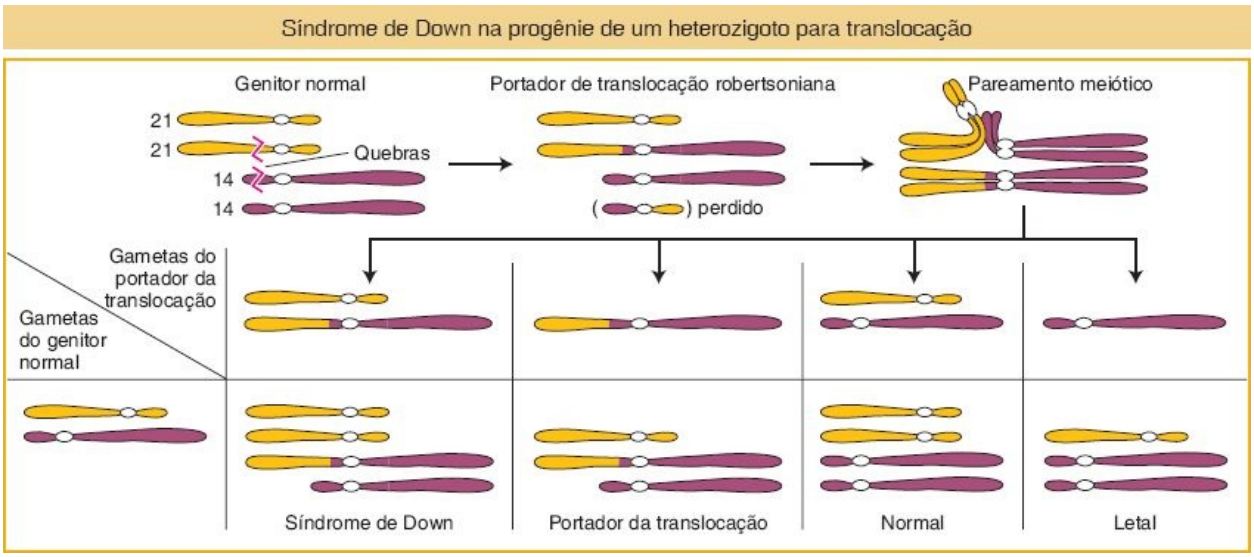


FIGURA 17.33 Na menor parte dos casos, a origem da síndrome de Down é um genitor heterozigoto para uma translocação robertsoniana envolvendo o cromossomo 21. A segregação meiótica resulta em alguns gametas que carregam um cromossomo com um grande segmento adicional do cromossomo 21. Em combinação com um cromossomo 21 normal fornecido pelo gameta do sexo oposto, são produzidos os sintomas da síndrome de Down, muito embora não exista uma trissomia completa do cromossomo 21.

Aplicações de inversões e translocações

As inversões e as translocações comprovaram ser ferramentas genéticas úteis; seguem alguns exemplos de suas utilizações.

Mapeamento gênico. As inversões e as translocações são úteis para o mapeamento e o subsequente isolamento de genes específicos. O gene da

neurofibromatose humana foi isolado desse modo. A informação crítica surgiu a partir de pessoas que não apenas apresentavam a doença, como também carregavam translocações cromossômicas. Todas as translocações apresentavam um ponto de quebra em comum, em uma banda próxima do centrômero do cromossomo 17. Portanto, essa banda aparentava ser o *locus* do gene da neurofibromatose, que havia sido rompido pelo ponto de quebra da translocação. A análise subsequente demonstrou que os pontos de quebra do cromossomo 17 não estavam em posições idênticas; entretanto, tendo em vista que eles devem estar no gene, a variação de suas posições revelou o segmento do cromossomo que constituía o gene da neurofibromatose. O isolamento de fragmentos de DNA dessa região finalmente levou à recuperação do próprio gene.

Síntese de duplicações ou deleções específicas. As translocações e as inversões são utilizadas de modo rotineiro para deletar ou duplicar segmentos cromossômicos específicos. Relembre, por exemplo, que as inversões pericêntricas, bem como as translocações, dão origem a produtos de meiose que contêm uma duplicação *e* uma deleção (ver Figuras 17.29 e 17.30). Se o segmento duplicado ou deletado for muito pequeno, então os produtos meióticos de duplicação e deleção são equivalentes às duplicações ou deleções, respectivamente. As duplicações e deleções são úteis para uma diversidade de aplicações experimentais, incluindo o mapeamento de genes e a variação da dosagem gênica para o estudo da regulação, conforme observado nas seções precedentes.

Outra abordagem para a criação de duplicações utiliza translocações *insercionais* unidirecionais, nas quais um segmento de um cromossomo é removido e inserido em outro. Em um heterozigoto para translocação insercional, resulta uma duplicação se o cromossomo com a inserção segregar juntamente com a cópia normal.

Variegação por efeito de posição. Conforme vimos no Capítulo 12, a ação gênica pode ser bloqueada pela proximidade com as regiões cromossômicas densamente coradas denominadas *heterocromatina*. As translocações e as inversões podem

ser utilizadas para estudar esse efeito. Por exemplo, o *locus* para a cor de olho branco em *Drosophila* está próximo da extremidade do cromossomo X. Considere uma translocação na qual a extremidade de um cromossomo X que carrega w^+ é realocada para perto da região heterocromática, digamos, do cromossomo 4 (Figura 17.34 A, *seção superior*). A **variegação por efeito de posição** é observada em moscas que são heterozigotas para tal translocação. O cromossomo X normal em um referido heterozigoto carrega o alelo recessivo *w*. Espera-se que o fenótipo do olho seja vermelho, em virtude de o alelo do tipo selvagem ser dominante sobre *w*. Entretanto, nos referidos casos, o fenótipo observado é uma mistura variegada de facetas oculares vermelhas e brancas (Figura 17.34 B). Como podemos explicar as áreas brancas? O alelo w^+ nem sempre é expresso, tendo em vista que o limite da heterocromatina é um pouco variável: em algumas células, ele envolve e inativa o gene w^+, evitando, assim, a sua expressão e possibilitando, assim, a expressão de *w*. Se as posições dos alelos w^+ e *w* forem permutadas por um *crossover*, então a variegação por efeito de posição não será detectada (ver Figura 17.34 A, *seção inferior*).

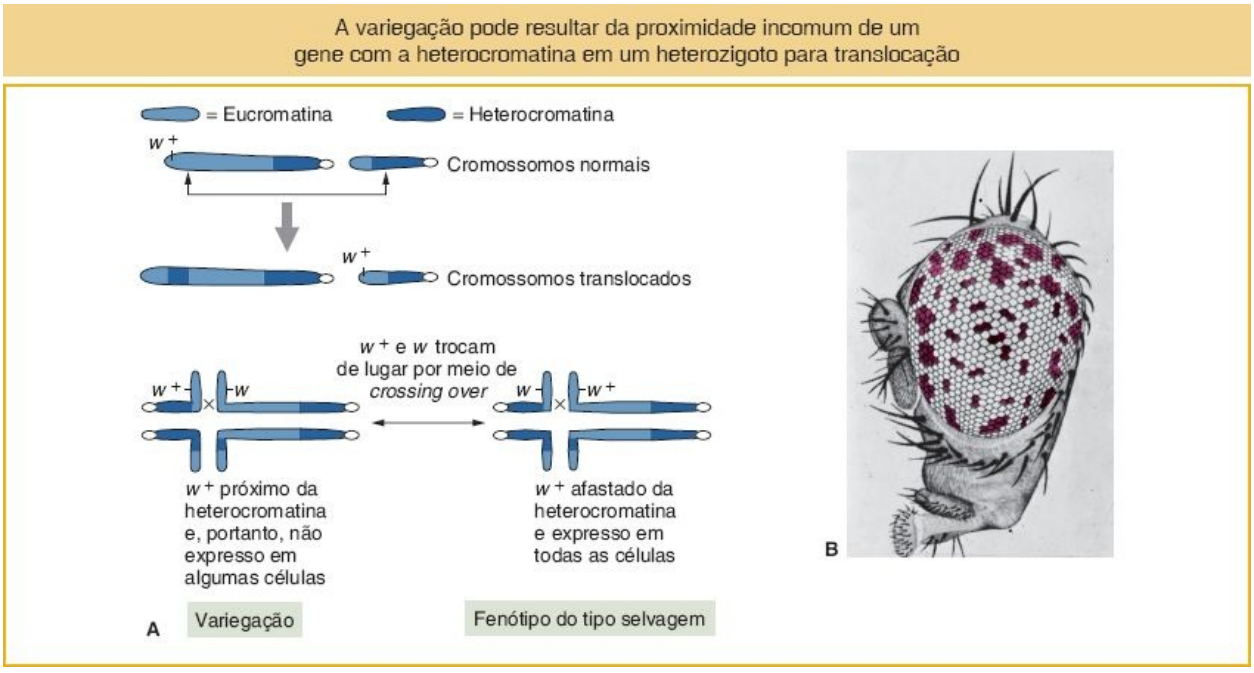


FIGURA 17.34 A. A translocação de w^+ para uma posição próxima da heterocromatina causa a falha da função de w^+ em algumas células, produzindo a variegação por efeito de posição. **B.** Um olho de *Drosophila*

que demonstra variegação por efeito de posição. (*B.* Gordon Watts.)

Rearranjos e câncer

O câncer é uma doença de proliferação celular anormal. Como resultado de alguma lesão, uma célula do corpo se divide descontroladamente até formar uma população de células denominada câncer. Um aglomerado localizado de células proliferadas é denominado tumor, enquanto os cânceres de células móveis, tais como as células sanguíneas, dispersam-se por todo o corpo. O câncer é causado com mais frequência por uma mutação na sequência codificadora ou reguladora de um gene cuja função normal é regular a divisão celular. Tais genes são denominados *proto-oncogenes*. Entretanto, os rearranjos cromossômicos, especialmente as translocações, também podem interferir na função normal de tais proto-oncogenes.

Existem dois modos básicos por meio dos quais as translocações podem alterar a função dos proto-oncogenes. No primeiro mecanismo, a translocação realoca um proto-oncogene para perto de um novo elemento regulador. Um bom exemplo é fornecido pelo linfoma de Burkitt. O proto-oncogene nesse câncer codifica a proteína MYC, um fator de transcrição que ativa genes necessários para a proliferação celular. Normalmente, o gene *myc* é transcrito apenas quando uma célula precisa ser submetida à proliferação, mas, em células cancerosas, o proto-oncogene *MYC* é realocado para perto da região reguladora dos genes de imunoglobulina (Ig) (Figura 17.35 A). Esses genes de imunoglobulina são transcritos constitutivamente; ou seja, eles estão ligados o tempo todo. Consequentemente, o gene *myc* é transcrito em todas as ocasiões e os genes de proliferação celular estão continuamente ativados.

O outro mecanismo por meio do qual as translocações podem causar câncer é a formação de um gene híbrido. Um exemplo é fornecido pela leucemia mieloide crônica (CML), um câncer dos leucócitos. Esse câncer pode resultar da formação de um gene híbrido entre os dois proto-oncogenes *BRC1* e *ABL* (Figura 17.35 B). O proto-oncogene *abl* codifica uma proteinoquinase em uma via de sinalização. A proteinoquinase transmite um sinal iniciado por um fator de crescimento, que leva à proliferação celular. A proteína de fusão de Brc1-Abl apresenta uma atividade

de proteinoquinase permanente. A proteína alterada propaga continuamente o seu sinal de crescimento adiante, independentemente da presença do sinal iniciador.

Identificação de mutações cromossômicas pela genômica

Os microarranjos de DNA (ver Figura 14.19) tornaram possível detectar e quantificar duplicações ou deleções de um determinado segmento de DNA. A técnica é denominada *hibridização genômica comparativa*. O DNA total do tipo selvagem e aquele de um mutante são marcados com dois corantes fluorescentes diferentes, que emitem comprimentos de onda de luz distintos. Esses DNA marcados são adicionados a um microarranjo de cDNA e ambos hibridizam com o arranjo. Em seguida, o arranjo é submetido a uma varredura por um detector ajustado para um comprimento de onda fluorescente e, depois, é novamente submetido a outra varredura, em outro comprimento de onda. É calculada a razão dos valores para cada cDNA. As razões de mutantes em relação ao tipo selvagem substancialmente superiores a 1 representam regiões que foram amplificadas. Uma razão de 2 pontos representa uma duplicação e uma razão inferior a 1 ponto, uma deleção. Alguns exemplos estão demonstrados na Figura 17.36.

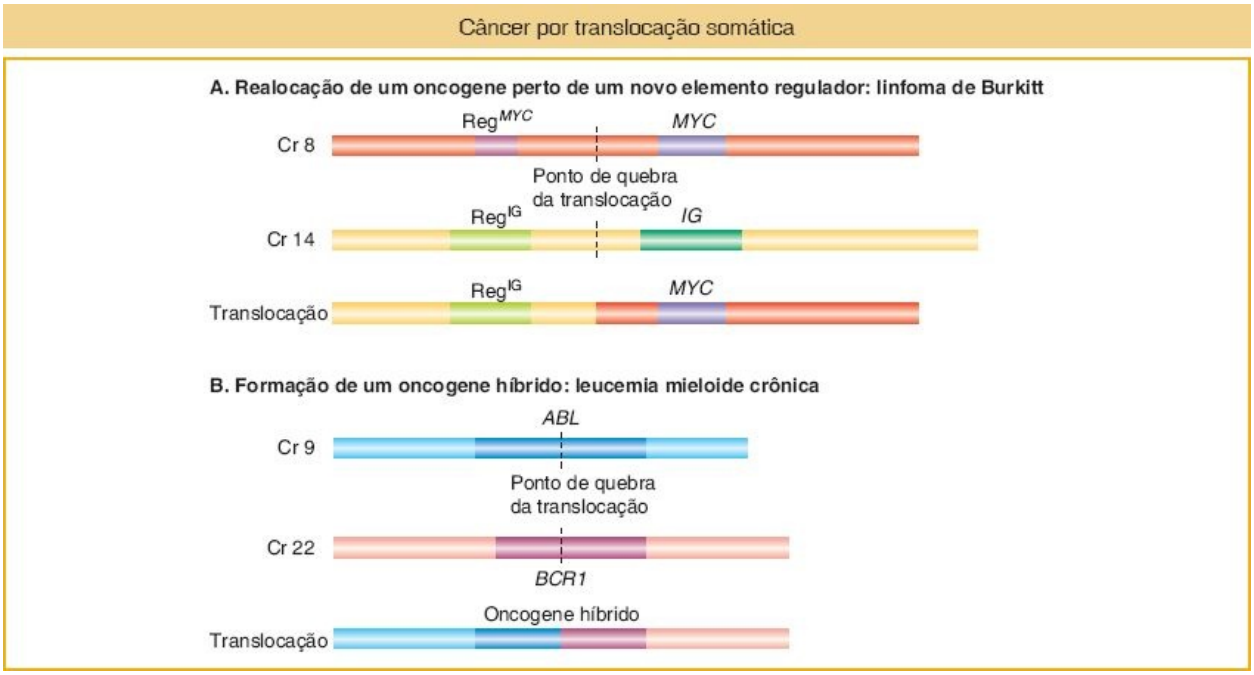


FIGURA 17.35 Os dois modos principais por meio dos quais as translocações podem causar câncer em uma

célula somática do corpo são ilustrados pelos cânceres linfoma de Burkitt (**A**) e leucemia mieloide crônica (**B**). Os genes *MYC*, *BRC1* e *ABL* são proto-oncogenes.

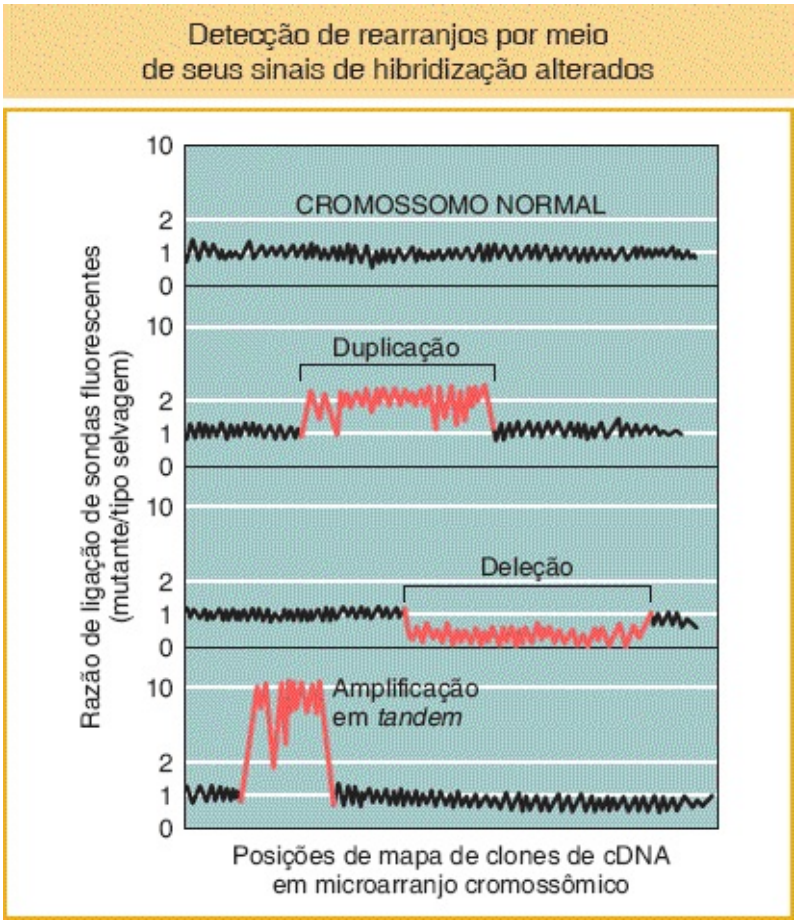


FIGURA 17.36 Para detectar rearranjos cromossômicos, o DNA genômico mutante e do tipo selvagem são marcados com corantes que fluorescem em diferentes comprimentos de onda. Estes DNA marcados são adicionados a clones de cDNA dispostos em microarranjos cromossomicamente ordenados e a razão de fluorescência ligada em cada comprimento de onda é calculada em relação a cada clone. Estão ilustrados os resultados esperados para um genoma normal e três tipos de mutantes.

17.3 Incidência geral de mutações cromossômicas humanas

As mutações cromossômicas surpreendentemente surgem com frequência na reprodução sexuada humana, demonstrando que os processos celulares relevantes são propensos a um alto nível de erros. A Figura 17.37 demonstra a distribuição

estimada de mutações cromossômicas entre as concepções humanas que se desenvolvem o suficiente para a implantação no útero. Dos estimados 15% de concepções que abortam espontaneamente (gestações que são encerradas naturalmente), metade demonstra anormalidades cromossômicas. Alguns geneticistas clínicos acreditam que até mesmo esse alto nível seja uma subestimativa, tendo em vista que muitos casos nunca são detectados. Entre os nascimentos vivos, 0,6% apresentam anormalidades cromossômicas, que resultam de aneuploidia e de rearranjos cromossômicos.

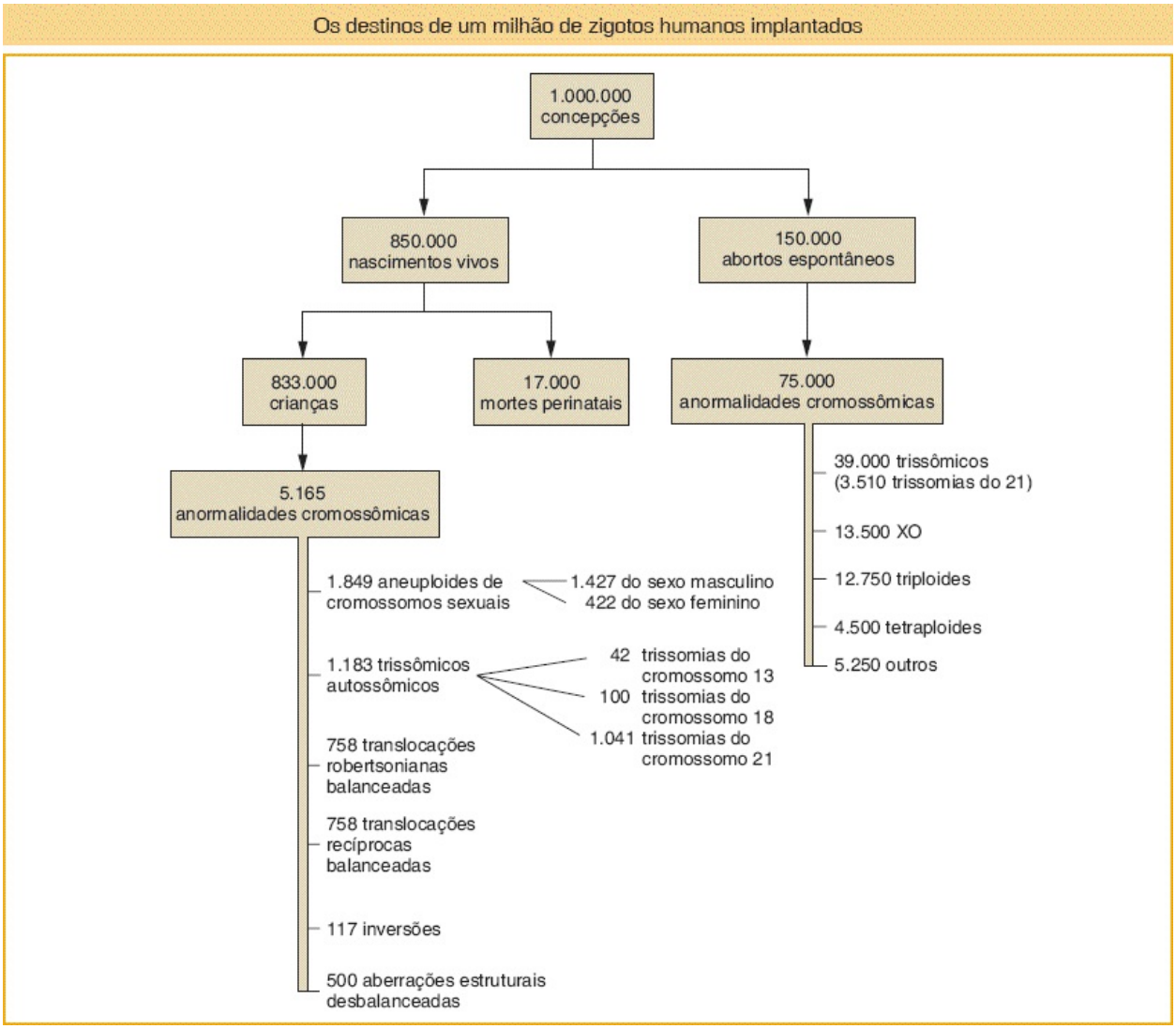


FIGURA 17.37 A proporção de mutações cromossômicas é muito mais alta em abortos espontâneos. (Dados de K. Sankaranarayanan, *Mutat. Res.* 61, 1979, 249-257.)

RESUMO

A poliploidia é uma condição anormal na qual existe um número de conjuntos cromossômicos superior ao normal. Os poliploides, tais como triploides (3*n*) e os tetraploides (4*n*), são comuns entre as plantas e estão representados também em animais. Os organismos com um número ímpar de conjuntos cromossômicos são estéreis, tendo em vista que nem todo cromossomo apresenta um parceiro na meiose. Os cromossomos não pareados se juntam aleatoriamente aos polos da célula na meiose, levando a conjuntos desbalanceados de cromossomos nos gametas resultantes. Tais gametas desbalanceados não produzem progênie viável. Em poliploides com um número par de conjuntos, cada cromossomo apresenta um parceiro de pareamento em potencial e, portanto, pode produzir gametas balanceados e progênie. A poliploidia pode resultar em um organismo de dimensões maiores; essa descoberta possibilitou avanços importantes na horticultura e nas plantações comerciais.

Em plantas, os alopoliploides (poliploides formados pela combinação de conjuntos cromossômicos de diferentes espécies) podem ser produzidos por meio do cruzamento de duas espécies relacionadas e, em seguida, da duplicação dos cromossomos da progênie por meio da utilização da colchicina ou por meio da fusão de células somáticas. Essas técnicas apresentam aplicações potenciais na agricultura, tendo em vista que os alopoliploides combinam as características das duas espécies parentais.

Quando acidentes celulares alteram partes dos conjuntos cromossômicos, resultam os aneuploides. A própria aneuploidia normalmente resulta em um genótipo desbalanceado com um fenótipo anormal. Exemplos de aneuploides incluem monossômicos (2*n* — 1) e trissômicos (2*n* + 1). A síndrome de Down (trissomia do cromossomo 21), a síndrome de Klinefelter (XXY) e a síndrome de Turner (XO) são exemplos bem-documentados de condições aneuploides em seres humanos. O nível espontâneo de aneuploidia em seres humanos é consideravelmente alto e é responsável por uma grande proporção de enfermidades com base genética em populações humanas. O fenótipo de um organismo aneuploide depende muito do cromossomo afetado em particular. Em

alguns casos, como na trissomia do cromossomo 21 humano, existe uma constelação altamente característica de fenótipos associados.

A maior parte dos casos de aneuploidia resulta da segregação cromossômica errônea acidental na meiose (não disjunção). O erro é espontâneo e pode ocorrer em qualquer meiócito particular na primeira ou na segunda divisão. Em seres humanos, um efeito da idade materna está associado à não disjunção do cromossomo 21, resultando em uma mais alta incidência de síndrome de Down nos filhos de mães mais velhas.

A outra categoria geral de mutações cromossômicas abrange os rearranjos estruturais, que incluem deleções, duplicações, inversões e translocações. Essas alterações resultam da quebra e da reunião incorreta ou de *crossing over* entre elementos repetitivos (recombinação homóloga não alélica). Os rearranjos cromossômicos são uma causa importante de enfermidades em populações humanas e são úteis na construção de linhagens especiais de organismos para a genética experimental e aplicada. Em organismos com um conjunto cromossômico normal mais um conjunto rearranjado (rearranjos heterozigotos), existem estruturas de pareamento incomuns na meiose, que resultam da forte afinidade de pareamento das regiões cromossômicas homólogas. Por exemplo, as inversões heterozigotas demonstram alças, e as translocações recíprocas demonstram estruturas em forma de cruz. A segregação dessas estruturas resulta em produtos meióticos anormais únicos do rearranjo.

Uma deleção é a perda de uma parte do cromossomo, seja em virtude de quebras cromossômicas seguidas pela perda do segmento interveniente, seja em virtude da segregação em translocações ou inversões heterozigotas. Se a região removida em uma deleção for essencial à vida, uma deleção homozigota é letal. As deleções heterozigotas podem ser letais em virtude do desequilíbrio cromossômico ou em virtude de revelarem alelos deletérios recessivos, ou podem ser não letais. Quando uma deleção em um homólogo possibilita a expressão fenotípica de alelos recessivos no outro, o desmascaramento dos alelos recessivos é denominado pseudodominância.

As duplicações em geral são produzidas a partir de outros rearranjos ou por meio de *crossing over* aberrante. Elas também desequilibram o material genético,

produzindo um efeito fenotípico deletério ou a morte do organismo. Entretanto, as duplicações podem ser uma fonte de novo material para a evolução, tendo em vista que a função pode ser mantida em uma cópia, deixando a outra cópia livre para desenvolver novas funções.

Uma inversão é um giro de 180° de parte de um cromossomo. No estado homozigoto, as inversões podem causar poucos problemas para um organismo, exceto se a heterocromatina ocasionar um efeito de posição ou se uma das quebras romper um gene. Por outro lado, os heterozigotos para inversão demonstram alças de inversão na meiose, e o *crossing over* dentro da alça resulta em produtos inviáveis. Os produtos do *crossover* de inversões pericêntricas, que abrangem o centrômero, diferem daqueles das inversões paracêntricas, que não abrangem, mas ambos demonstram redução da frequência de recombinantes na região afetada e com frequência resultam em redução da fertilidade.

Uma translocação movimenta um segmento cromossômico para outra posição no genoma. Um exemplo simples é uma translocação recíproca, na qual partes de cromossomos não homólogos trocam de posições. No estado heterozigoto, as translocações geram produtos meióticos com duplicação e deleção, que podem levar a zigotos desbalanceados. Novas ligações gênicas podem ser produzidas por meio de translocações. A segregação aleatória dos centrômeros em um heterozigoto com translocação resulta em 50% de produtos meióticos desbalanceados e, portanto, 50% de esterilidade (semiesterilidade).

TERMOS-CHAVE

alça de deleção
alça de inversão
alopoliploide
aneuploide
anfidiploide
autopoliploide
balanceador
balanço gênico

bivalente
carga genética
compensação de dose
cromossomo acêntrico
cromossomo dicêntrico
cromossomo homeólogo
cromossomo politênico
deleção
deleção intragênica
deleção multigênica
dissômico
duplicação
duplicação em *tandem*
duplicação insercional
duplicação segmentar
efeito de dosagem gênica
embrioide
euploide
fragmento acêntrico
heterozigoto para inversão
hexaploide
inversão
inversão paracêntrica
inversão pericêntrica
mapeamento de deleção
monoploide
monossômico
mutação cromossômica
não disjunção
nulissômico
partenogênese
pentaploide

poliploide
ponte anafásica
ponte dicêntrica
pseudodominância
pseudoligação
rearranjo
rearranjo balanceado
rearranjo desbalanceado
recombinação homóloga não alélica (NAHR)
segregação adjacente 1
segregação alternada
semiesterilidade
síndrome de Down
síndrome de Klinefelter
síndrome de Turner
tetraploide
translocação
triploide
trissômico
trivalente
univalente
variegação por efeito de posição

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Uma planta de milho é heterozigota em relação a uma translocação recíproca e, portanto, é semiestéril. Essa planta é cruzada com uma linhagem cromossomicamente normal que é homozigota em relação ao alelo recessivo braquítico (*b*), localizado no cromossomo 2. Uma planta da F_1 semiestéril em seguida é retrocruzada com a linhagem braquítica homozigota. A progênie obtida demonstra os fenótipos a seguir:

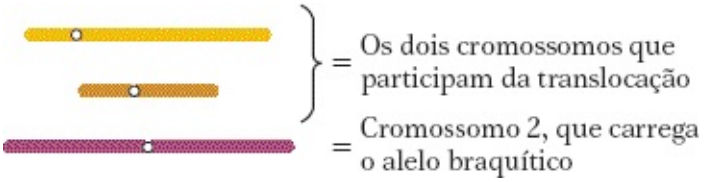
Não braquítica		*Braquítica*	
Semiestéril	Fértil	Semiestéril	Fértil
334	27	42	279

a. Qual razão você espera que resulte se o cromossomo que carrega o alelo braquítico não participar da translocação?

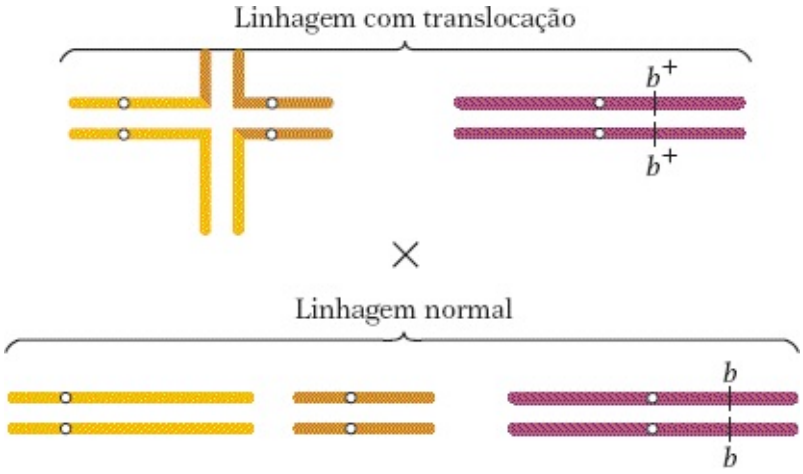
b. Você acredita que o cromossomo 2 participa da translocação? Explique a sua resposta, demonstrando a conformação dos cromossomos relevantes da F_1 semiestéril e o motivo dos números específicos obtidos.

Solução

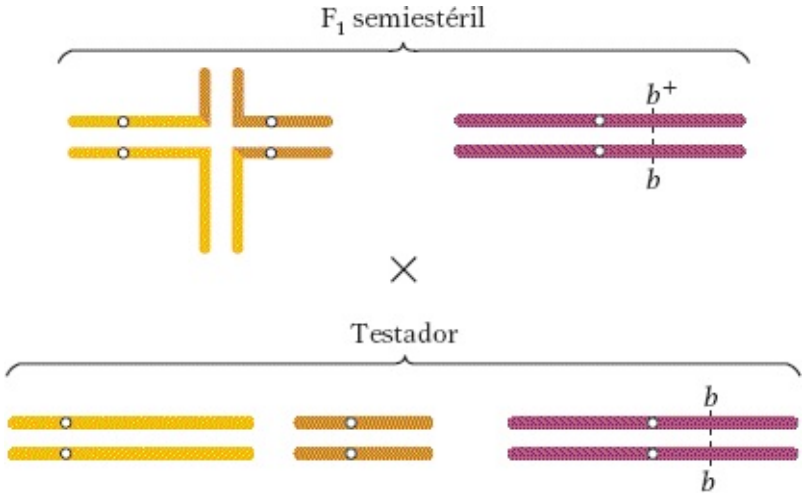
a. Devemos iniciar com a abordagem metódica e simplesmente reformular os dados como um diagrama, no qual:



Para simplificar o diagrama, não demonstramos os cromossomos divididos em cromátides (embora eles estejam nesse estágio da meiose). Em seguida, diagramamos o primeiro cruzamento:



Toda a progênie desse cruzamento será heterozigota em relação ao cromossomo que carrega o alelo braquítico, mas o que dizer sobre os cromossomos que participam da translocação? Neste capítulo, vimos que apenas os produtos da segregação alternada sobrevivem e que metade desses sobreviventes será cromossomicamente normal, além de metade carregar os dois cromossomos rearranjados. A combinação rearranjada regenerará um heterozigoto com translocação quando se combinar com o complemento cromossomicamente normal do genitor normal. Esses últimos tipos — os semiestéreis da F_1 — são diagramados como parte do retrocruzamento com a linhagem braquítica parental:



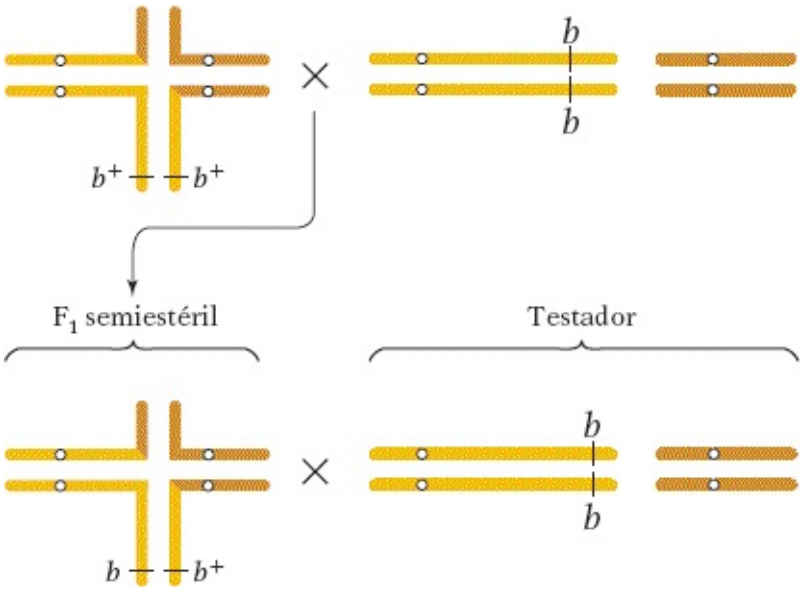
No cálculo da razão esperada de fenótipos a partir desse cruzamento, podemos tratar o comportamento dos cromossomos translocados independentemente do comportamento do cromossomo 2. Portanto, podemos prever que a progênie será:

$\frac{1}{2}$ heterozigoto para translocação (semiestéril)
- $\frac{1}{2}$ b^+/b → $\frac{1}{4}$ não braquítico semiestéril
- $\frac{1}{2}$ b/b → $\frac{1}{4}$ braquítico semiestéril

$\frac{1}{2}$ normal (fértil)
- $\frac{1}{2}$ b^+/b → $\frac{1}{4}$ não braquítico fértil
- $\frac{1}{2}$ b/b → $\frac{1}{4}$ braquítico fértil

Essa razão de 1:1:1:1 prevista é bem diferente daquela obtida no cruzamento real.

b. Em virtude de observarmos um desvio da razão esperada com base na independência do fenótipo braquítico e da semiesterilidade, o cromossomo 2 provavelmente *participa* da translocação. Presumiremos que o *locus* braquítico (*b*) está no cromossomo laranja. Mas, onde? Para a finalidade do diagrama, não importa onde o posicionamos, mas isso importa geneticamente, tendo em vista que a posição do *locus b* afeta as razões na progênie. Se presumirmos que o *locus b* está próximo da extremidade do pedaço que é translocado, podemos redesenhar o heredograma:



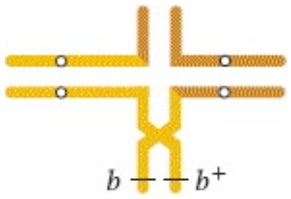
Se os cromossomos da F_1 semiestéril segregarem como diagramado aqui, podemos então prever:

$\frac{1}{2}$ fértil, braquítico

$\frac{1}{2}$ semiestéril, não braquítico

A maior parte da progênie certamente é desse tipo e, assim, devemos estar no rumo certo. Como os dois tipos menos frequentes são produzidos? De algum modo, precisamos posicionar o alelo b^+ no cromossomo amarelo normal e o alelo *b* no cromossomo translocado. Esse posicionamento deve ser conquistado por meio de *crossing over* entre o ponto de quebra da translocação (o centro da

estrutura em forma de cruz) e o *locus* braquítico:



Os cromossomos recombinantes produzem alguma progênie que é fértil e não braquítica e alguma que é semiestéril e braquítica (essas duas classes em conjunto constituem 69 de uma progênie total de 682 ou uma frequência de aproximadamente 10%). Podemos observar que essa frequência é realmente uma medida da distância de mapa (10 u.m.) do *locus* braquítico ao ponto de quebra. (O mesmo princípio básico teria sido obtido se houvéssemos desenhado o *locus* braquítico na parte do cromossomo no outro lado do ponto de quebra.)

Problema resolvido 2. Possuímos linhagens de camundongos puras para dois fenótipos comportamentais alternativos que sabemos serem determinados por dois alelos em um único *locus*: *v* faz um camundongo se movimentar em marcha de "valsador", enquanto *V* determina marcha normal. Após o cruzamento dos valsadores puros com os normais, observamos que a maior parte da F_1 é normal, mas, inesperadamente, existe uma fêmea valsadora. Cruzamos a valsadora da F_1 com dois machos valsadores diferentes e observamos que ela produz apenas progênie valsadora. Quando a cruzamos com machos normais, ela produz progênie normal e nenhum valsador. Cruzamos três fêmeas de sua progênie normal com dois de seus irmãos, e esses camundongos produzem uma progênie de 60 camundongos, todos normais. Entretanto, quando cruzamos uma dessas mesmas três fêmeas com um terceiro irmão, obtemos seis normais e dois valsadores em uma ninhada de oito. Ao pensar a respeito dos genitores da valsadora da F_1, podemos considerar algumas possíveis explicações desses resultados:

a. Um alelo dominante pode ter sofrido mutação para um alelo recessivo em seu genitor normal.

b. Em um genitor, pode ter havido uma mutação dominante em um segundo gene para criar um alelo epistático, que atua para prevenir a expressão de *V*, levando à

característica valsadora.

c. A não disjunção meiótica do cromossomo que carrega *V* no seu genitor normal pode ter dado origem a um aneuploide viável.

d. Pode ter havido uma deleção viável abrangendo *V* no meiócito de seu genitor normal.

Quais dessas explicações são possíveis e quais são eliminadas pela análise genética? Explique detalhadamente.

Solução

O melhor modo de responder à questão é considerar as explicações, uma por vez, e verificar se cada uma delas se encaixa nos resultados fornecidos.

a. Mutação de *V* para *v*

Essa hipótese requer que a fêmea valsadora excepcional seja homozigota *v/v*. Essa presunção é compatível tanto com os resultados do seu cruzamento com machos valsadores, com os quais, se ela for *v/v*, produziria toda a progênie valsadora (*v/v*), quanto com os machos normais, com que produziria toda a progênie normal (*V/v*). Entretanto, os cruzamentos de irmão e irmã dentro dessa progênie normal então produziriam uma razão de 3:1 de normais para valsadores. Tendo em vista que de fato alguns dos cruzamentos de irmão e irmã não produziram valsadores, essa hipótese não explica os dados.

b. Mutação epistática de *s* para *S*

Aqui, os genitores seriam *V/V · s/s* e *v/v · s/s*, e uma mutação germinativa em um deles forneceria à valsadora da F_1 o genótipo *V/v · S/s*. Quando realizássemos o seu cruzamento com um macho valsador, que seria do genótipo *v/v · s/s*, esperaríamos alguma progênie *V/v · S/s*, que seria fenotipicamente normal. Entretanto, não observamos progênie normal a partir desse cruzamento e, assim, a hipótese já é derrubada. A ligação poderia salvar a hipótese temporariamente se presumíssemos que a mutação ocorreu no genitor normal, fornecendo um gameta *VS*. Então, a valsadora da F_1 seria *VS/vs* e, se a ligação fosse suficientemente forte, poucos gametas *Vs* seriam produzidos, ou mesmo nenhum, tipo que é necessário para combinar com o gameta *vs* do macho para originar os normais *Vs/vs*. Entretanto, se a hipótese de ligação fosse verdadeira, o cruzamento com os

machos normais seria *VS/vs* × *Vs/Vs*, e isso proporcionaria uma alta porcentagem de progênie *VS/Vs*, que seria valsadora; nenhum valsador foi observado.

c. Não disjunção no genitor normal

Essa explicação forneceria um gameta nulissômico, que se combinaria com *v* para originar a valsadora da F_1 com o genótipo hemizigoto *v*. Os cruzamentos subsequentes seriam:

- *v* × *v/v*, que fornece progênie *v/v* e *v*, todos valsadores. Essa se encaixa
- *v* × *V/v*, que fornece progênie *V/v* e *V*, todos normais. Essa também se encaixa
- Primeiro intercruzamento de progênie normal: *V* × *V*. Esse intercruzamento fornece *V* e *V/V*, que são normais. Essa se encaixa
- Segundo intercruzamento de progênie normal: *V* × *V/v*. Esse intercruzamento fornece 25% cada de *V/V, V/v, V* (todas normais), e *v* (valsadores). Essa também se encaixa.

Portanto, essa hipótese é consistente com os dados.

d. Deleção de *V* no genitor normal

Denominaremos a deleção D. A valsadora da F_1 seria D/*v*, e os cruzamentos subsequentes seriam:

- D/*v* × *v/v*, que fornece *v/v* e D/*v*, valsadores. Essa se encaixa
- D/*v* × *V/V*, que fornece *V/v* e D/*V*, normais. Essa se encaixa
- Primeiro intercruzamento de progênie normal: D/*V* × D/*V*, que fornece D/*V* e *V/V*, todos normais. Essa se encaixa
- Segundo intercruzamento de progênie normal: D/*V* × *V/v*, que fornece 25% cada de *V/V, V/v*, D/*V* (todos normais), e D/*v* (valsadores). Essa também se encaixa.

Mais uma vez, a hipótese se encaixa nos dados fornecidos; assim, restam-nos duas hipóteses que são compatíveis com os resultados, e são necessários experimentos adicionais para distingui-las. Um modo de fazer isso seria examinar os cromossomos da fêmea excepcional ao microscópio; a aneuploidia seria fácil de distinguir da deleção.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Com base na Tabela 17.1, como você categorizaria os genomas a seguir? (As letras H a J fazem referência a quatro cromossomos diferentes.)

HH II J KK
HH II JJ KKK
HHHH IIII JJJJ KKKK

2. Com base na Figura 17.4, quantas cromátides estão em um trivalente?
3. Com base na Figura 17.5, se a colchicina for utilizada em uma planta na qual $2n = 18$, quantos cromossomos estariam no produto anormal?
4. Baseando o seu trabalho na Figura 17.7, utilize canetas coloridas para representar os cromossomos do anfidiploide fértil.
5. Se o trigo de Emmer (Figura 17.9) fosse cruzado com outro trigo selvagem CC (não demonstrado), qual seria a constituição de um produto estéril desse cruzamento? Qual anfidiploide poderia surgir a partir do produto estéril? O anfidiploide seria fértil?
6. Na Figura 17.12, qual seria a constituição de um indivíduo formado a partir da união de um monossômico resultante de uma não disjunção na primeira divisão em uma fêmea e um dissômico resultante de uma não disjunção de segunda divisão em um macho, presumindo que os gametas sejam funcionais?
7. Na Figura 17.14, qual seria a porcentagem esperada de cada tipo de segregação?
8. Na Figura 17.19, existe qualquer diferença entre os produtos de inversão formados a partir da quebra e aqueles formados a partir de *crossing over*?
9. Fazendo referência à Figura 17.19, desenhe um diagrama demonstrando o processo por meio do qual uma inversão formada a partir de *crossing over* poderia gerar uma sequência normal.
10. Na Figura 17.21, um alelo *fa* recessivo seria expresso quando pareado com a deleção 264-32? E com 265-11?

11. Observe a Figura 17.22 e declare quais bandas são perdidas na deleção de *cri du chat*.

12. Na Figura 17.25, qual espécie está relacionada de modo mais próximo com a linhagem de levedura ancestral? Por que os genes 3 e 13 são considerados duplicados?

13. Fazendo referência à Figura 17.26, desenhe o produto se ocorreram quebras nos genes *A* e *B*.

14. Na Figura 17.26, o painel inferior demonstra que os genes *B* e *C* estão orientados em uma direção diferente (observe os promotores). Você acredita que essa diferença na orientação afetaria a sua funcionalidade?

15. Na Figura 17.28, qual seria a consequência de um *crossover* entre o centrômero e o *locus* A?

16. Com base na Figura 17.30, genomas normais chegam a ser formados a partir dos dois tipos de segregação? Genomas normais chegam a ser formados a partir de uma segregação adjacente 1?

17. Fazendo referência à Figura 17.32, desenhe um produto *inviável* a partir da mesma meiose.

18. Com base na Figura 17.35, escreva uma sentença declarando como a translocação pode levar ao câncer. Você consegue pensar a respeito de outra causa genética de câncer?

19. Observando a Figura 17.36, por que você acha que a razão do sinal é tão mais alta no painel inferior?

20. Utilizando a Figura 17.37, calcule qual porcentagem das concepções é de triploides. A mesma figura demonstra XO na categoria de aborto espontâneo; entretanto, sabemos que muitos indivíduos XO são viáveis. Em qual das categorias de viáveis os XO seriam agrupados?

PROBLEMAS BÁSICOS

21. Ao manter o estilo da Tabela 17.1, como você denominaria os organismos

que são MM N OO; MM NN OO; MMM NN PP?

22. Uma grande planta surgiu em uma população natural. Qualitativamente, ela aparentava ser justamente como as outras, exceto por ser muito maior. É mais provável que ela seja um alopoliploide ou um autopoliploide? Como você testaria se ela é um poliploide, não apenas uma planta crescida em um solo rico?

23. Um trissômico é um aneuploide ou um poliploide?

24. Em um tetraploide *B/B/b/b*, existem quantos pareamentos quadrivalentes possíveis? Desenhe-os (ver Figura 17.5).

25. Alguém lhe diz que a couve-flor é um anfidiploide. Você concorda? Explique.

26. Por que *Raphanobrassica* é fértil, enquanto a seu genitor não era?

27. Na designação dos genomas de trigo, quantos cromossomos são representados pela letra B?

28. Como você "recriaria" o trigo hexaploide a partir de *Triticum tauschii* e Emmer?

29. Como você produziria uma muda monoploide ao iniciar com uma planta diploide?

30. É obtido um produto dissômico da meiose. Qual é a sua provável origem? Quais outros genótipos você esperaria entre os produtos daquela meiose sob a sua hipótese?

31. Um trissômico *A/A/a* pode chegar a produzir um gameta de genótipo *a*?

32. Quais, se houver algum, dos aneuploides de cromossomos sexuais em seres humanos a seguir são férteis: XXX, XXY, XYY, XO?

33. Por que as mães gestantes mais velhas realizam amniocentese ou obtenção e análise de amostra de vilosidade coriônica como procedimento de rotina?

34. Em uma inversão, uma extremidade de 5′ do DNA chega a ser unida a outra extremidade de 5′? Explique.

35. Se você observou uma ponte dicêntrica na meiose, qual rearranjo você prevê que tenha ocorrido?

36. Por que os fragmentos acêntricos são perdidos?

37. Diagrame uma translocação que surge a partir de DNA repetitivo. Repita em relação a uma deleção.

38. A partir de um grande estoque de rearranjos de *Neurospora* disponíveis do centro de estoques genéticos de fungos, qual tipo você escolheria para sintetizar uma linhagem apresentando uma duplicação do braço direito do cromossomo 3 e uma deleção da extremidade do cromossomo 4?

39. Você observa uma alça de pareamento muito grande na meiose. É mais provável que ela seja de uma inversão heterozigota ou de uma deleção heterozigota? Explique.

40. Um novo alelo mutante recessivo não demonstra pseudodominância com quaisquer das deleções que abrangem o cromossomo 2 de *Drosophila*. Qual pode ser a explicação?

41. Compare e contraste as origens da síndrome de Turner, da síndrome de Williams, da síndrome *cri du chat* e da síndrome de Down. (Por que elas são denominadas *síndromes*?)

42. Liste as características diagnósticas (genéticas ou citológicas) que são utilizadas para identificar essas alterações cromossômicas:

a. Deleções

b. Duplicações

c. Inversões

d. Translocações recíprocas

43. A sequência normal de nove genes em um determinado cromossomo de *Drosophila* é 123 · 456789, na qual o ponto representa o centrômero. Observou-se que algumas moscas-das-frutas apresentam cromossomos aberrantes com as estruturas a seguir:

a. 123 · 476589

b. 123 · 46789

c. 1654 · 32789

d. 123 · 4566789

Denomine cada tipo de rearranjo cromossômico, e desenhe diagramas para demonstrar como cada um realizaria sinapse com o cromossomo normal.

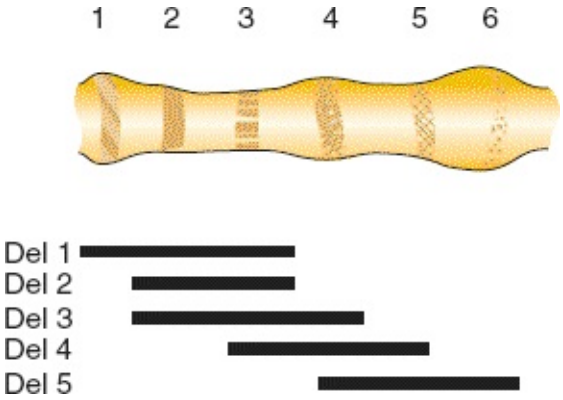
44. Os dois *loci P* e *Bz* normalmente estão à distância de 36 u.m. no mesmo braço de um determinado cromossomo de planta. Uma inversão paracêntrica abrange aproximadamente um quarto dessa região, mas não inclui qualquer um dos *loci*. Qual frequência de recombinantes aproximada entre *P* e *Bz* você preveria em plantas que são:

a. Heterozigotas em relação à inversão paracêntrica?

b. Homozigotas em relação à inversão paracêntrica?

45. Conforme declarado no Problema resolvido 2, a mutação recessiva em determinados camundongos denominados *valsadores* faz com que eles executem passos bizarros. W. H. Gates cruzou valsadores com camundongos normais de raça pura e observou, entre diversas centenas de progênie normal, uma única fêmea valsadora. Ela foi cruzada com um macho valsador, e toda sua progênie foi valsadora. Quando cruzada com um macho normal homozigoto, toda a progênie foi normal. Alguns desses machos e fêmeas normais foram intercruzados e, inesperadamente, não houve descendentes valsadores. T. S. Painter examinou os cromossomos de alguns dos camundongos valsadores de Gates que demonstravam um comportamento reprodutivo semelhante àquele da fêmea valsadora incomum original. Ele observou que esses camundongos apresentavam o número normal de 40 cromossomos. Entretanto, nos valsadores incomuns, um membro de um par de cromossomos era anormalmente curto. Interprete essas observações tão completamente quanto possível, tanto genética quanto citologicamente.

46. Um cromossomo da glândula salivar de *Drosophila* apresenta seis bandas, conforme demonstrado na ilustração a seguir. Abaixo do cromossomo estão demonstradas as extensões de cinco deleções (Del 1 a Del 5):



Os alelos recessivos *a*, *b*, *c*, *d*, *e* e *f* sabidamente estão na região, mas a sua ordem é desconhecida. Quando as deleções são combinadas com cada alelo, são obtidos os resultados a seguir:

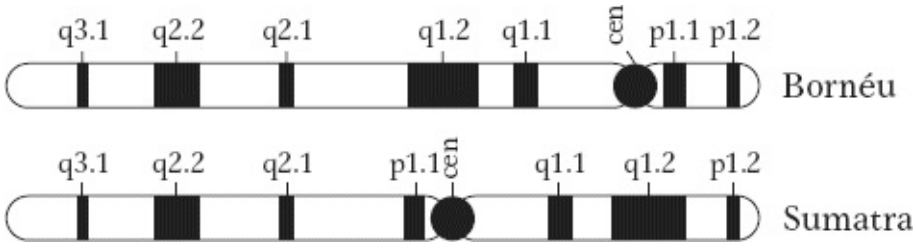
	a	*b*	*c*	*d*	*e*	*f*
Del 1	—	—	—	+	+	+
Del 2	—	+	—	+	+	+
Del 3	—	+	—	+	—	+
Del 4	+	+	—	—	—	+
Del 5	+	+	+	—	—	—

Nesta tabela, um sinal de menos significa que a deleção revela o alelo recessivo (é observado o fenótipo recessivo), e um sinal de mais significa que o alelo do tipo selvagem correspondente ainda está presente. Corresponda cada banda salivar com um gene.

47. Observou-se que uma mosca-das-frutas é heterozigota em relação a uma

inversão paracêntrica. Entretanto, a obtenção de moscas que eram homozigotas em relação à inversão foi impossível, até mesmo após muitas tentativas. Qual é a explicação mais provável para essa incapacidade de produzir uma inversão homozigota?

48. Orangotangos são uma espécie ameaçada em seu ambiente natural (as ilhas de Bornéu e Sumatra) e, assim, foi estabelecido um programa de criação em cativeiro com a utilização de orangotangos atualmente mantidos em zoológicos por todo o mundo. Um componente desse programa é pesquisar a citogenética dos orangotangos. Essa pesquisa demonstrou que todos os orangotangos de Bornéu carregam um tipo de cromossomo 2, conforme demonstrado no diagrama que acompanha, e todos os orangotangos de Sumatra carregam o outro tipo. Antes de essa diferença citogenética se tornar conhecida, foram realizados alguns cruzamentos entre animais de diferentes ilhas e 14 progênies híbridas atualmente estão sendo criadas em cativeiro.



a. Qual termo ou quais termos, descreve(m) as diferenças entre estes cromossomos?

b. Desenhe o cromossomo 2, pareado na primeira prófase meiótica, de tal orangotango híbrido. Assegure-se de demonstrar todos os pontos de referência indicados no diagrama que acompanha, e rotule todas as partes de seu desenho.

c. Em 30% das meioses, haverá um *crossover* em algum local na região entre as bandas p1.1 e q1.2. Desenhe os cromossomos 2 do gameta que resultaria de uma meiose na qual ocorreu um único *crossover* dentro da banda q1.1.

d. Qual fração dos gametas produzidos por um orangotango híbrido dará origem a progênie viável, se esses cromossomos forem os únicos que diferem entre os genitores? (O Problema 48 é de Rosemary Redfield.)

49. No milho, os genes para o comprimento da borla (alelos *T* e *t*) e para resistência à ferrugem (alelos *R* e *r*) sabidamente estão em cromossomos separados. No período da realização de cruzamentos de rotina, um cultivador observou que uma planta *T/t; R/r* proporcionava resultados incomuns em um cruzamento-teste com o pólen de um genitor duplo recessivo *t/t; r/r*. Os resultados foram:

Progênies:	*T/t; R/r*	98
	t/t; r/r	104
	T/t; r/r	3
	t/t; R/r	5
Espigas de milho:	apenas aproximadamente metade com muitos grãos como de costume	

a. Quais características-chave dos dados são diferentes dos resultados esperados?

b. Elabore uma hipótese concisa que explique os resultados.

c. Demonstre os genótipos dos genitores e da progênie.

d. Desenhe um diagrama demonstrando o arranjo de alelos nos cromossomos.

e. Explique a origem das duas classes de progênie que demonstram três e cinco membros.

Como solucionar o Problema 49

1. O que um "gene em relação ao comprimento "da borda" e um "gene em relação à resistência à ferrugem" significam?

2. É de importância que o significado preciso dos símbolos alélicos *T, t, R* e *r* não seja fornecido? Por que sim ou por que não?
3. Como os termos *gene* e *alelo*, conforme utilizados aqui, estão relacionados com os conceitos de *locus* e par de genes?
4. Qual evidência experimental anterior forneceria ao geneticista de milho a ideia de que dois genes estão em cromossomos separados?
5. O que você imagina que "cruzamentos de rotina" sejam para um cultivador de milho?
6. Qual termo é utilizado para descrever os genótipos do tipo *T/t; R/r*?
7. O que é um "pólen genitor"?
8. O que são cruzamentos-teste, e por que os geneticistas acham que eles são tão úteis?
9. Quais tipos de progênie e frequências o cultivador poderia estar esperando a partir do cruzamento-teste?
10. Descreva como a progênie observada difere das expectativas.
11. O que a igualdade aproximada das duas primeiras classes de progênie informa a você?
12. O que a igualdade aproximada das duas segundas classes de progênie informa a você?
13. Quais eram os gametas da planta incomum, e quais eram as suas proporções?
14. Quais gametas estavam na maioria?
15. Quais gametas estavam na minoria?
16. Quais dos tipos de progênie aparentam ser recombinantes?
17. Quais combinações alélicas aparentam estar ligadas de algum modo?
18. Como pode haver ligação de genes supostamente em cromossomos separados?
19. O que essas classes majoritárias e minoritárias nos informam a respeito dos genótipos dos genitores da planta incomum?
20. O que é uma espiga de milho?
21. Qual é o aspecto de uma espiga de milho normal? (Esboce uma e a rotule.)
22. Qual é o aspecto das espigas de milho desse cruzamento? (Esboce uma.)
23. O que exatamente é um grão de milho?

24. Qual efeito poderia levar à ausência de metade dos grãos de milho?
25. Metade dos grãos de milho morreu? Em caso afirmativo, o motivo das mortes foi o genitor do sexo feminino ou do masculino?

Agora tente solucionar o problema.

50. Corpo amarelo em *Drosophila* é causado por um alelo mutante *y* de um gene localizado na extremidade do cromossomo X (o alelo do tipo selvagem causa um corpo cinza). Em um experimento com radiação, um macho do tipo selvagem foi irradiado com raios X e em seguida cruzado com uma fêmea de corpo amarelo. A maior parte da progênie do sexo masculino era amarela, conforme esperado, mas a varredura de milhares de moscas revelou dois machos de corpos cinza (fenotipicamente do tipo selvagem). Esses machos de corpos cinza foram cruzados com fêmeas de corpos amarelos, com os resultados a seguir:

	Progênie
Macho cinza 1 × Fêmea amarela	Todas as fêmeas amarelas
	Todos os machos cinza
Macho cinza 2 × Fêmea amarela	$\frac{1}{2}$fêmeas cinza
fêmeas amarelas	
	$\frac{1}{2}$machos amarelos
	$\frac{1}{2}$machos cinza

a. Explique a origem e o comportamento de cruzamento do macho cinza 1.
b. Explique a origem e o comportamento de cruzamento do macho cinza 2.

51. No milho, o alelo *Pr* define talos verdes e *pr*, roxos. Uma planta do milho de genótipo *pr/pr* que apresenta cromossomos-padrão é cruzada com uma planta *Pr/Pr* que é homozigota em relação a uma translocação recíproca entre os cromossomos 2 e 5. A F_1 é semiestéril e fenotipicamente Pr. Um retrocruzamento com o genitor com cromossomos-padrão dá origem a 764 Pr semiestéreis, 145 pr semiestéreis, 186 Pr normais e 727 pr normais. Qual é a distância de mapa entre o *locus Pr* e o ponto de translocação?

52. Diferencie entre as síndromes de Klinefelter, Down e Turner. Quais síndromes são observadas em ambos os sexos?

53. Demonstre como você poderia produzir um alotetraploide entre duas espécies de plantas diploides relacionadas, ambas $2n = 28$.

54. Em *Drosophila,* trissômicos e monossômicos em relação ao pequeno cromossomo 4 são viáveis, mas nulissômicos e tetrassômicos não são. O *locus b* está nesse cromossomo. Deduza as proporções fenotípicas na progênie dos cruzamentos de trissômicos a seguir.

a. $b^+/b/b \times b/b$

b. $b^+/b^+/b \times b/b$

c. $b^+/b^+/b \times b^+/b$

55. Observa-se que uma mulher com síndrome de Turner é daltônica (um fenótipo recessivo ligado ao X). Tanto sua mãe quanto seu pai apresentam visão normal.

a. Explique a origem simultânea da síndrome de Turner e do daltonismo por meio do comportamento anormal dos cromossomos na meiose.

b. A sua explicação pode distinguir se o comportamento cromossômico anormal ocorreu no pai ou na mãe?

c. A sua explicação pode distinguir se o comportamento cromossômico anormal ocorreu na primeira ou na segunda divisão da meiose?

d. Agora presuma que um homem com Klinefelter daltônico possui genitores com visão normal, e resposta as partes *a*, *b* e *c*.

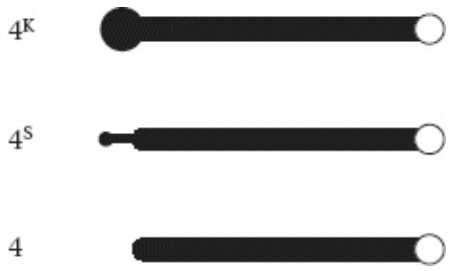
56. **a.** Como você sintetizaria um pentaploide?

b. Como você sintetizaria um triploide de genótipo *A/a/a*?

c. Você acabou de obter uma mutação recessiva rara *a** em uma planta diploide, que a análise mendeliana informa a você ser *A/a**. A partir dessa planta, como você sintetizaria um tetraploide (4*n*) de genótipo *A/A/a*/a**?

d. Como você sintetizaria um tetraploide de genótipo *A/a/a/a*?

57. Suponha que você possui uma linhagem de camundongos que apresenta tipos citologicamente distintos do cromossomo 4. A extremidade do cromossomo pode apresentar uma protuberância (denominado 4^K) ou um satélite (4^S) ou nenhum deles (4). Aqui estão os esboços dos três tipos:



Você cruza uma fêmea $4^K/4^S$ com um macho 4/4 e observa que a maior parte da progênie é $4^K/4$ ou $4^S/4$, conforme o esperado. Entretanto, ocasionalmente você observa alguns tipos raros, como segue (todos os outros cromossomos são normais):

a. $4^K/4^K/4$

b. $4^K/4^S/4$

c. 4^K

Explique os tipos raros que você observou. Forneça, tão precisamente quanto possível, os estágios nos quais eles se originam e declare se eles são originados no genitor, na genitora ou no zigoto. (Forneça brevemente os motivos.)

58. É realizado um cruzamento em tomates entre uma planta do sexo feminino que é trissômica em relação ao cromossomo 6 e uma planta do sexo masculino diploide normal que é homozigota em relação ao alelo recessivo relativo à folha de *potato* (*p/p*). Uma planta F_1 trissômica é submetida ao retrocruzamento com o macho de folhas de *potato*.

a. Qual é a razão de plantas com folhas normais e plantas com folhas de

potato quando você presume que *p* está localizado no cromossomo 6?

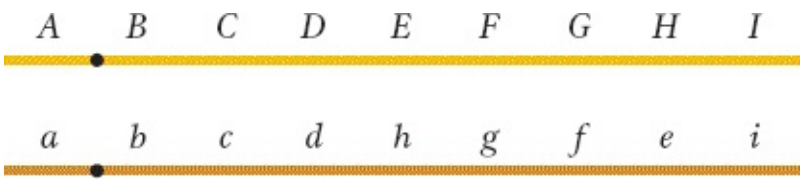
b. Qual é a razão de plantas com folhas normais e plantas com folhas de *potato* quando você presume que *p* não está localizado no cromossomo 6?

59. Uma geneticista de tomate tenta atribuir cinco mutações recessivas a cromossomos específicos por meio da utilização de trissômicos. Ela cruza cada mutante homozigoto (2*n*) com cada um de três trissômicos, nos quais os cromossomos 1, 7 e 10 participam. A partir desses cruzamentos, a geneticista seleciona progênie trissômica (que é menos vigorosa) e realiza o seu retrocruzamento com o homozigoto recessivo apropriado. É examinada a progênie *diploide* a partir desses cruzamentos. Os seus resultados, nos quais as razões são tipo selvagem:mutante, são como segue:

Cromossomo trissômico	*Mutação*				
	d	*y*	*c*	*h*	*cot*
1	48:55	72:29	56:50	53:54	32:28
7	52:56	52:48	52:51	58:56	81:40
10	45:42	36:33	28:32	96:50	20:17

Quais das mutações a geneticista pode atribuir a quais cromossomos? (Explique totalmente a sua resposta.)

60. Uma petúnia é heterozigota em relação aos homólogos autossômicos a seguir:



a. Desenhe a configuração de pareamento que você observaria na metáfase I, e identifique todas as partes do seu diagrama. Numere as cromátides sequencialmente a partir da parte superior até a parte inferior da página.

b. Ocorre um *crossover* duplo de três filamentos, com um *crossover* entre os *loci C* e *D* nas cromátides 1 e 3, e o segundo *crossover* entre os *loci G* e *H* nas cromátides 2 e 3. Diagrame os resultados desses eventos de recombinação, conforme você os observaria na anáfase I, e identifique todas as partes do seu diagrama.

c. Desenhe o padrão cromossômico que você observaria na anáfase II após os *crossovers* descritos na parte *b*.

d. Forneça os genótipos dos gametas dessa meiose que levariam à formação de progênie viável. Presuma que todos os gametas são fertilizados pelo pólen que apresenta a ordem de genes *A B C D E F G H I*.

61. Dois grupos de geneticistas, na Califórnia e no Chile, começam a trabalhar para desenvolver um mapa de ligação de *medfly* (mosca-das-frutas do Mediterrâneo). Independentemente, ambos descobrem que os *loci* em relação à cor do corpo (*B* = Preto; *b* = Cinza) e ao formato dos olhos (*R* = Redondos, *r* = Estrelados) estão ligados com uma distância de 28 u.m. Eles enviam linhagens uns aos outros e realizam cruzamentos; um resumo de todos os seus achados está demonstrado aqui:

Cruzamento	F_1	Progênie de F_1 × qualquer *br/br*	
		BR/br	36%
		br/br	36

BR/BR (Calif.) × *br/br* (Calif.)	*BR/br*	*Br/br*	14
		bR/br	14
BR/BR (Chile) × *br/br* (Chile)	*BR/br*	*BR/br*	36
		br/br	36
		Br/br	14
		bR/br	14
BR/BR (Calif.)× *br/br* (Chile) ou *br/br* (Calif.) × *BR/BR* (Chile)	*BR/br*	*BR/br*	48
		br/br	48
		Br/br	2
		bR/br	2

a. Forneça uma hipótese genética que explique os três conjuntos de resultados dos cruzamentos-teste.

b. Desenhe as características cromossômicas-chave da meiose na F_1 a partir de um cruzamento das linhagens da Califórnia e do Chile.

62. Uma planta do milho aberrante proporciona os valores de FR a seguir quando submetida a um cruzamento-teste:

	Intervalo

	d-f	*f-b*	*b-x*	*x-y*	*y-p*
Controle	5	18	23	12	6
Planta aberrante	5	2	2	0	6

(A ordem do *locus* é centrômero-*d-f-b-x-y-p*.) A planta aberrante é uma planta saudável, mas produz muito menos óvulos normais e pólen do que a planta-controle produz.

a. Proponha uma hipótese para explicar os valores de recombinação anormais e a redução da fertilidade na planta aberrante.

b. Utilize diagramas para explicar a origem dos recombinantes de acordo com a sua hipótese.

63. Os *loci* do milho a seguir estão em um braço do cromossomo 9, na ordem indicada (as distâncias entre eles estão demonstradas em unidades de mapa):

c—*bz*—*wx*—*sh*—*d*—centrômero
12 8 10 20 10

C proporciona aleurona colorida; *c,* aleurona branca;

Bz proporciona folhas verdes; *bz,* folhas bronze.

Wx proporciona sementes amiláceas; *wx,* sementes cerosas.

Sh proporciona sementes lisas; *sh,* sementes enrugadas.

D proporciona plantas altas; *d,* anãs.

Uma planta de um estoque padrão que é homozigota em relação a todos os cinco alelos recessivos é cruzada com uma planta do tipo selvagem do México, que é homozigota em relação a todos os cinco alelos dominantes. As plantas da F_1 expressam todos os alelos dominantes e, quando retrocruzadas com o genitor recessivo, proporciona os fenótipos de progênie a seguir:

colorida, verde, amilácea, lisa, alta	360
branca, bronze, cerosa, enrugada, anã	355
colorida, bronze, cerosa, enrugada, anã	40
branca, verde, amilácea, lisa, alta	46
colorida, verde, amilácea, lisa, anã	85
branca, bronze, cerosa, enrugada, alta	84
colorida, bronze, cerosa, enrugada, alta	8
branca, verde, amilácea, lisa, anã	9
colorida, verde, cerosa, lisa, alta	7
branca, bronze, amilácea, enrugada, anã	6

Proponha uma hipótese para explicar esses resultados. Inclua:
a. Uma declaração geral da sua hipótese, com diagramas, se necessário
b. O motivo da existência de 10 classes
c. Uma explicação da origem de cada classe, incluindo sua frequência e
d. No mínimo um teste de sua hipótese.

64. Plantas de milho cromossomicamente normais apresentam um *locus p* no cromossomo 1 e um *locus s* no cromossomo 5.

P proporciona folhas verde-escuras; *p,* folhas verde-pálidas.

S proporciona espigas grandes; *s*, espigas murchas.

Uma planta original de genótipo *P/p; S/s* apresenta o fenótipo esperado (verde-escura, espigas grandes), mas proporciona resultados inesperados em cruzamentos, como segue:

- Na autofecundação, a fertilidade é normal, mas a frequência de tipos *p/p*; *s/s* é de 1/4 (não de 1/16, conforme o esperado)
- Quando cruzadas com um testador normal de genótipo *p/p*; *s/s*, a progênie da F_1 é 1/2 *P/p*; *S/s* e 1/2 *p/p*; *s/s*; a fertilidade é normal
- Quando uma planta da F_1 *P/p*; *S/s* é cruzada com um testador normal *p/p*; *s/s*, ela comprova ser semiestéril, mas, novamente, a progênie é 1/2 *P/p*; *S/s* e 1/2 *p/p*; *s/s*.

Explique esses resultados, demonstrando os genótipos completos da planta original, do testador, e das plantas da F_1. Como você testaria a sua hipótese?

	Embriões (número médio)			
Cruzamento	Implantados na parede uterina	Degeneração após a implantação	Normais	Degeneração (%)
♂ excepcional × ♀ normal	8,7	5,0	3,7	37,5
♂ normal × ♀ normal	9,5	0,6	8,9	6,5

65. Um rato macho que é fenotipicamente normal demonstra anomalias

reprodutivas quando comparado com ratos machos normais, conforme demonstrado na tabela anterior. Proponha uma explicação genética para esses resultados incomuns e indique como a sua ideia poderia ser testada.

66. Uma geneticista de tomate que trabalha com *Fr*, um alelo mutante dominante que causa o rápido amadurecimento do fruto, decide descobrir qual cromossomo contém esse gene por meio da utilização de um conjunto de linhagens, das quais cada uma é trissômica em relação a um cromossomo. Para tanto, ela cruza um mutante diploide homozigoto com cada uma das linhagens trissômicas do tipo selvagem.

a. Uma planta trissômica da F_1 é cruzada com uma planta do tipo selvagem diploide. Qual é a razão das plantas de amadurecimento rápido e lento na progênie diploide desse segundo cruzamento se o *Fr* está no cromossomo trissômico? Utilize diagramas para explicar.

b. Qual é a razão de plantas de amadurecimento rápido e lento na progênie diploide desse segundo cruzamento se *Fr* não está localizado no cromossomo trissômico? Utilize diagramas para explicar.

c. Aqui estão os resultados dos cruzamentos. Em qual cromossomo está *Fr*, e por quê?

Cromossomo trissômico	Amadurecimento rápido:amadurecimento lento na progênie diploide
1	45:47
2	33:34
3	55:52
4	26:30

5	31:32
6	37:41
7	44:79
8	49:53
9	34:34
10	37:39

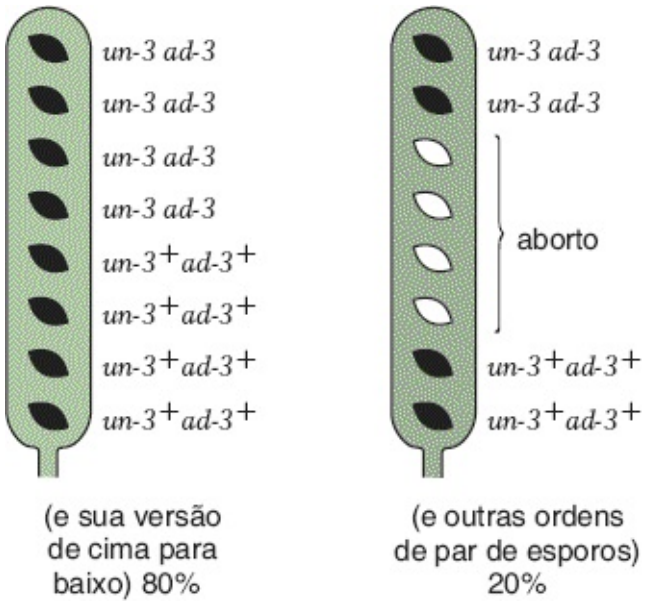
(O Problema 66 é de Tamara Western.)

PROBLEMAS DESAFIADORES

67. O *locus un-3* de *Neurospora* está próximo do centrômero no cromossomo 1, e *crossovers* entre *un-3* e o centrômero são muito raros. O *locus ad-3* está do outro lado do centrômero do mesmo cromossomo, e ocorrem *crossovers* entre *ad-3* e o centrômero em aproximadamente 20% das meioses (não ocorrem *crossovers* múltiplos).

a. Quais tipos de ascos lineares (ver Capítulo 4) você prevê, com que frequência, em um cruzamento normal de *un-3 ad-3* × tipo selvagem? (Especifique os genótipos dos esporos nos ascos.)

b. Na maior parte do tempo, tais cruzamentos se comportam de modo previsível, mas, em um caso, uma linhagem *un-3 ad-3* padrão foi cruzada com um tipo selvagem isolado em um campo de cana-de-açúcar no Havaí. Seguem os resultados:



Explique esses resultados e declare como você poderia testar a sua ideia. (Nota: em *Neurospora*, os ascósporos com material cromossômico extra sobrevivem e são da cor preta normal, enquanto os ascósporos com ausência de qualquer região cromossômica são brancos e inviáveis.)

68. Duas mutações em *Neurospora*, *ad-3* e *pan-2*, estão localizadas nos cromossomos 1 e 6, respectivamente. Uma linhagem *ad-3* incomum surge em laboratório, fornecendo os resultados demonstrados na tabela a seguir. Explique todos os três resultados com o auxílio de diagramas claramente marcados. (Nota: em *Neurospora*, os ascósporos com material cromossômico extra sobrevivem e são da cor preta normal, enquanto os ascósporos com ausência de qualquer região cromossômica são brancos e inviáveis.)

	Aspecto dos ascósporos	FR entre *ad-3* e *pan-2*
1. *ad-3* normal × *pan-2* normal	Todos pretos	50%

2. *ad-3* anormal × *pan-2* normal	Aproximadamente $\frac{1}{2}$ pretos $\frac{1}{2}$ e brancos (inviáveis)	1%

3. Dos esporos pretos do cruzamento 2, aproximadamente metade era completamente normal e metade repetiu o mesmo comportamento da linhagem *ad-3* anormal original

---

69. Deduza as proporções fenotípicas na progênie dos cruzamentos a seguir de autotetraploides, nos quais os *locus* a^+/a está muito próximo do centrômero. (Presuma que os quatro cromossomos homólogos de qualquer tipo pareiam aleatoriamente, dois a dois, e que apenas uma cópia do alelo a^+ é necessária para o fenótipo do tipo selvagem.)

a. $a^+/a^+/a/a \times a/a/a/a$

b. $a^+/a/a/a \times a/a/a/a$

c. $a^+/a/a/a \times a^+/a/a/a$

d. $a^+/a^+/a/a \times a^+/a/a/a$

70. A espécie de algodão do Novo Mundo *Gossypium hirsutum* apresenta um número 2*n* de 52 cromossomos. As espécies do Velho Mundo *G. thurberi* e *G. herbaceum* apresentam, cada uma, um número 2*n* de 26. Quando essas espécies são cruzadas, os híbridos resultantes demonstram os arranjos de pareamento cromossômico na meiose a seguir:

Híbrido	Arranjo de pareamento
G. hirsutum × *G. thurberi*	13 bivalentes pequenos + 13 univalentes grandes
G. hirsutum × *G. herbaceum*	13 bivalentes grandes + 13 univalentes pequenos

G. thurberi × *G. herbaceum*	13 univalentes grandes + 13 univalentes pequenos

Interprete essas observações filogeneticamente, utilizando diagramas. Indique claramente as relações entre as espécies. Como você comprovaria que a sua interpretação está correta?

71. Existem seis espécies principais do gênero *Brassica*: *B. carinata, B. campestris, B. nigra, B. oleracea, B. juncea* e *B. napus*. Você consegue deduzir as inter-relações dessas seis espécies a partir da tabela a seguir:

a. Deduza o número de cromossomos de *B. campestris, B. nigra* e *B. oleracea.*

b. Demonstre claramente quaisquer relações evolutivas entre as seis espécies que você consiga deduzir no nível cromossômico.

Espécie ou híbrido de F_1	Número de cromossomos	Número de bivalentes	Número de univalentes
B. juncea	36	18	0
B. carinata	34	17	0
B. napus	38	19	0
B. juncea × *B. nigra*	26	8	10
B. napus × *B. campestris*	29	10	9

B. carinata × *B. oleracea*	26	9	8
B. juncea × *B. oleracea*	27	0	27
B. carinata × *B. campestris*	27	0	27
B. napus × *B. nigra*	27	0	27

---

72. Diversos tipos de mosaicismo sexual estão bem-documentados em seres humanos. Sugira como cada um dos exemplos a seguir pode ter surgido por meio da não disjunção na *mitose*:

a. XX/XO (ou seja, existem dois tipos celulares no corpo, XX e XO)

b. XX/XXYY

c. XO/XXX

d. XX/XY

e. XO/XX/XXX

73. Em *Drosophila*, foi realizado um cruzamento (cruzamento 1) entre duas moscas mutantes, uma homozigota em relação à mutação recessiva asas dobradas (*b*) e a outra homozigota em relação à mutação recessiva sem olhos (*e*). As mutações *e* e *b* são alelos de dois genes diferentes que sabidamente estão ligados de modo muito próximo no pequeno cromossomo 4 autossômico. Toda a progênie apresenta um fenótipo do tipo selvagem. Uma fêmea da progênie foi cruzada com um macho de genótipo *be/be*; denominaremos esse cruzamento 2. A maior parte da progênie do cruzamento 2 foi dos tipos esperados, mas houve também uma fêmea rara de fenótipo do tipo selvagem.

a. Explique que progênie comum espera-se do cruzamento 2.
b. A fêmea do tipo selvagem rara pode ter surgido por meio de (1) *crossing over* ou (2) não disjunção? Explique.
c. A fêmea do tipo selvagem rara foi submetida ao cruzamento-teste com um macho de genótipo *be/be* (cruzamento 3). A progênie foi:

$\frac{1}{6}$ tipo selvagem

$\frac{1}{3}$ asas dobradas

$\frac{1}{3}$ asas dobradas, sem olhos

$\frac{1}{2}$ sem olhos

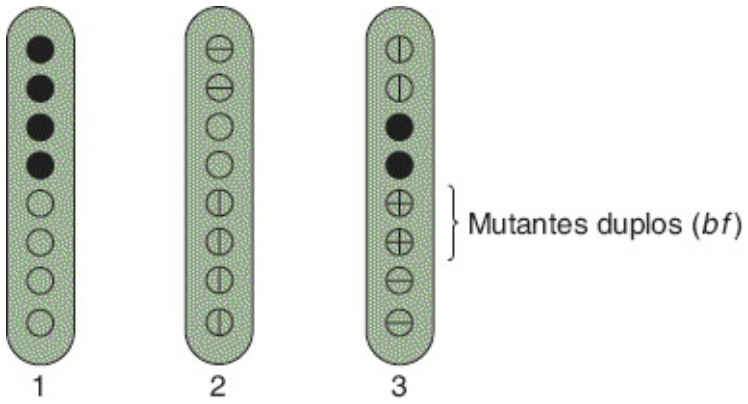
Qual das explicações na parte *b* é compatível com esse resultado? Explique os genótipos e os fenótipos da progênie do cruzamento 3 e as suas proporções.

Como solucionar o Problema 73

1. Defina *homozigoto, mutação, alelo, estreitamente ligado, recessivo, tipo selvagem, crossing over, não disjunção, cruzamento-teste, fenótipo* e *genótipo.*
2. Esse problema diz respeito à ligação ao sexo? Explique.
3. Quantos cromossomos a *Drosophila* apresenta?
4. Desenhe um heredograma claro resumindo os resultados dos cruzamentos 1, 2 e 3.
5. Desenhe os gametas produzidos por ambos os genitores no cruzamento 1.
6. Desenhe a constituição do cromossomo 4 da progênie do cruzamento 1.
7. É surpreendente que a progênie do cruzamento 1 seja de fenótipo do tipo selvagem? O que esse desfecho informa a você?
8. Desenhe a constituição do cromossomo 4 do testador macho utilizado no cruzamento 2 e os gametas que ele pode produzir.

9. Em relação ao cromossomo 4, quais gametas o genitor feminino no cruzamento 2 pode produzir na ausência de não disjunção? Quais seriam comuns e quais seriam raros?
10. Desenhe a não disjunção meiótica de primeira e de segunda divisão no genitor feminino do cruzamento 2, bem como nos gametas resultantes.
11. Algum dos gametas da parte 10 é aneuploide?
12. Você esperaria que os gametas aneuploides dessem origem a progênie viável? Essa progênie seria nulissômica, monossômica, dissômica ou trissômica?
13. Quais fenótipos de progênie seriam produzidos pelos diversos gametas considerados nas partes 9 e 10?
14. Considere a razão fenotípica na progênie do cruzamento 3. Muitas razões genéticas têm por base as metades e os quartos, mas essa razão tem por base os terços e os sextos. Essa razão pode apontar para o quê?
15. Pode haver qualquer significância no fato de que os cruzamentos dizem respeito a genes em um cromossomo muito pequeno? Quando o tamanho do cromossomo é relevante em genética?
16. Desenhe a progênie esperada a partir do cruzamento 3 sob as duas hipóteses, e forneça alguma ideia das proporções relativas.

74. No fungo *Ascobolus* (semelhante a *Neurospora*), os ascósporos normalmente são pretos. A mutação *f*, que produz ascósporos de cor castanho-clara, está em um gene logo à direita do centrômero no cromossomo 6, enquanto a mutação *b*, que produz ascósporos bege, está em um gene logo à esquerda do mesmo centrômero. Em um cruzamento de genitores castanho-claros e bege ($+f \times b+$), a maior parte das óctades demonstrou quatro ascósporos castanho-claros e quatro bege, mas foram observadas três óctades excepcionais raras, conforme demonstrado na ilustração a seguir. No esboço, o preto é o fenótipo do tipo selvagem, uma linha vertical é castanho-clara, uma linha horizontal é bege e um círculo vazio representa um ascósporo abortado (morto).



a. Forneça explicações razoáveis para essas três óctades excepcionais.

b. Diagrame a meiose que deu origem à óctade 2.

75. O ciclo de vida do fungo haploide *Ascobolus* é semelhante àquele de *Neurospora*. Um tratamento mutacional produziu duas linhagens mutantes, 1 e 2, ambas cruzadas com o tipo selvagem, dando origem a tétrades desordenadas, todas do tipo a seguir (castanho-claro é uma cor marrom-clara; normalmente, os cruzamentos produzem ascósporos todos pretos):

par 1 de esporos pretos	par 3 de esporos castanho-claros
par 2 de esporos pretos	par 4 de esporos castanho-claros

a. O que esse resultado demonstra? Explique.

As duas linhagens mutantes foram cruzadas. A maior parte das tétrades desordenadas era do tipo a seguir:

par 1 de esporos castanho-claros	par 3 de esporos claros castanho-claros
par 2 de esporos castanho-claros	par 4 de esporos claros castanho-claros

b. O que esse resultado sugere? Explique.

Quando grandes números de tétrades desordenadas foram triados ao microscópio, foram observadas algumas tétrades raras que continham esporos pretos. Quatro casos estão demonstrados aqui:

	Caso A	Caso B	Caso C	Caso D
Par 1 de esporos	Preto	Preto	Preto	Preto
Par 2 de esporos	Preto	Castanho-claro	Preto	Abortado
Par 3 de esporos	Castanho-claro	Castanho-claro	Abortado	Castanho-claro
Par 4 de esporos	Castanho-claro	Castanho-claro	Abortado	Castanho-claro

(Nota: os ascósporos com material genético extra sobrevivem, mas aqueles com menos de um genoma haploide abortam.)

c. Proponha explicações genéticas razoáveis para cada um desses quatro casos raros.

d. Você acredita que as mutações nas duas linhagens mutantes originais estavam em um único gene? Explique.

CAPÍTULO

18

Genética de Populações

Concepção da artista Lynn Fellman do “Adão Eurasiano”, um homem africano com um cromossomo Y pertencente a um grupo de haplótipos que foi o ancestral de todos os cromossomos Y de homens fora da África e que surgiu na África há aproximadamente 70.000 anos. (Lynn Fellman, www.Fellmanstudio.com.)

TÓPICOS

18.1 Detecção da variação genética
18.2 Conceito de pool gênico e lei de Hardy-Weinberg
18.3 Sistemas de acasalamento
18.4 Variação genética e sua medida
18.5 Modulação da variação genética
18.6 Aplicações biológicas e sociais

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Analisar dados para determinar quanta variação genética existe nas populações
- Desenhar um experimento para testar se uma população de organismos está de acordo com as expectativas de Hardy-Weinberg
- Explicar como novos alelos entram em uma população
- Compreender os impactos negativos do endocruzamento em uma população
- Descrever os tipos de seleção em relação à genética de populações
- Prever de que maneira forças como a seleção, a mutação e a deriva genética alteram a quantidade de variação nas populações.

Em 2009, Sean Hodgson foi libertado de uma prisão britânica após cumprir 27 anos pelo assassinato de Teresa De Simone, balconista e garçonete de um bar. Hodgson, que sofre de doença mental, inicialmente confessou o crime, mas retirou a sua confissão durante o processo. Durante seus anos na prisão, ele afirmou a sua inocência. Mais de duas décadas após o crime, a justiça analisou o DNA do assaltante, encontrado na cena do crime, e determinou que ele não era de Hodgson. A sua condenação foi anulada e a polícia reabriu a investigação do assassinato da Sra. De Simone. Conforme você aprenderá neste capítulo, a análise com base no DNA utilizada para absolver o Sr. Hodgson e centenas de outros prisioneiros erroneamente condenados depende da análise

genética de populações.

Os princípios da genética de populações são o cerne de muitas questões que a sociedade enfrenta atualmente. Quais são os riscos de que um casal tenha um filho com uma doença genética? As práticas de cultivo de plantas e criação de animais causaram uma perda da diversidade genética na agroindústria que imponha riscos para o nosso suprimento alimentar? Na medida em que a população humana continua a se expandir e a vida selvagem recua para partes cada vez menores da Terra, as espécies selvagens serão capazes de evitar o endocruzamento e sobreviver? Os princípios da genética de populações também são fundamentais para a compreensão de muitas questões históricas e evolutivas. Como as populações humanas de diferentes regiões do mundo estão relacionadas umas às outras? Como o genoma humano reagiu à medida que os seres humanos se espalharam por todo o globo e se tornaram adaptados a diferentes ambientes e estilos de vida? Como as populações e as espécies evoluem ao longo do tempo?

Uma **população** é um grupo de indivíduos da mesma espécie. A **genética de populações** analisa a quantidade e a distribuição da variação genética nas populações e as forças que controlam essa variação. Ela tem suas origens no início do século 20, quando geneticistas começaram a estudar como as leis de Mendel poderiam ser estendidas para compreender a variação genética dentro de populações inteiras de organismos. Embora as leis de Mendel expliquem como os genes são transmitidos dos genitores para a descendência nos casos de cruzamentos controlados e heredogramas conhecidos, essas leis são insuficientes para a compreensão da transmissão de genes de uma geração para a próxima em populações naturais, nas quais nem todos os indivíduos produzem descendência e nem toda a descendência sobrevive. Geneticistas iniciaram o desenvolvimento dos princípios da genética de populações no início do século 20, mas naquela época eles tinham ferramentas um tanto quanto limitadas para realmente medir a variação genética. Com o desenvolvimento das tecnologias com base no DNA ao longo das últimas três décadas, os geneticistas atualmente têm a capacidade de observar diretamente as diferenças entre as sequências de DNA dos indivíduos em todo o seu genoma e podem medir essas diferenças em grandes amostras de indivíduos em muitas espécies. O resultado tem sido uma revolução na nossa

compreensão sobre a variação genética nas populações.

Neste capítulo, consideraremos o conceito de *pool* gênico e como os geneticistas estimam as frequências de alelos e de genótipos nas populações. Em seguida, examinaremos o impacto que os sistemas de cruzamento apresentam sobre as frequências de genótipos em uma população. Também discutiremos como os geneticistas medem a variação com a utilização de tecnologias baseadas no DNA. Em seguida, discutiremos as forças que modulam os níveis de variação genética nas populações. Finalmente, observaremos alguns estudos de caso que envolvem a aplicação da genética de populações para questões de interesse para a sociedade.

18.1 Detecção da variação genética

Os métodos da genética de populações podem ser utilizados para analisar qualquer variável ou *locus* polimórfico nas sequências de DNA de uma população de organismos. Historicamente, os geneticistas não tinham as ferramentas moleculares necessárias para observar diretamente as diferenças nas sequências de DNA entre os indivíduos e, assim, a maior parte das análises da genética de populações observava as diferenças nas proteínas ou nos fenótipos. Por exemplo, as diferenças na proteína glicosiltransferase codificada pelo gene *AB0*, que controla o grupo sanguíneo AB0 em seres humanos, podem ser detectadas com a utilização de sondas de anticorpos. A partir dessas diferenças na proteína, os investigadores podem inferir diferenças na sequência de DNA desse gene entre os indivíduos. Ao longo das últimas três décadas, foram desenvolvidas novas tecnologias, tais como sequenciamento do DNA, microarranjos de DNA e PCR (ver Capítulos 10 e 14), possibilitando que os geneticistas observem as diferenças nas sequências de DNA diretamente. Como resultado, as análises da genética de populações deixaram de estar confinadas a um pequeno conjunto de genes, tais como *AB0*, tendo sido expandidas para incluir todos os nucleotídios no genoma.

Na genética de populações, um ***locus*** é simplesmente um local no genoma; ele

pode ser um sítio de nucleotídio único ou um segmento de muitos nucleotídios. O tipo mais simples de variação que se pode observar entre os indivíduos em um *locus* é uma diferença no nucleotídio presente em um sítio de nucleotídio único, seja adenina, citosina, guanina ou timina. Esses tipos de variantes são denominados **polimorfismos de nucleotídio único (SNP)** e são as variantes mais amplamente estudadas na genética de populações humanas (Figura 18.1; ver também Capítulo 4). A genética de populações também faz uso extensivo dos *loci* de **microssatélites** (ver Capítulo 4). Esses *loci* apresentam um *motif* de sequência curta, com comprimento de dois a seis pares de bases, que é repetido múltiplas vezes com diferentes alelos e números diferentes de repetições. Por exemplo, o *motif* de sequência AG de 2 pb em um *locus* pode ser repetido em *tandem* cinco vezes em um alelo (AGAGAGAGAG), mas três vezes em outro (AGAGAG) (ver Figura 18.1).

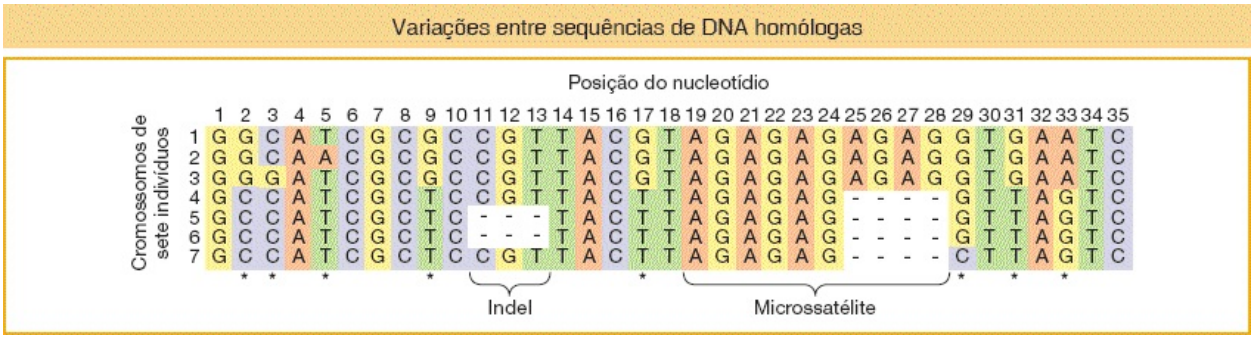


FIGURA 18.1 Variação nas sequências de DNA alinhadas de sete cromossomos de diferentes pessoas. Os *asteriscos* demonstram a localização dos SNP. A localização de uma indel (inserção/deleção de uma série de pares de nucleotídios) e de um microssatélite também está indicada.

Polimorfismos de nucleotídio único

Os polimorfismos de nucleotídio único (SNP) são os tipos de polimorfismo mais prevalentes na maior parte dos genomas. A maioria dos SNP apresenta apenas dois alelos — por exemplo, A e C. Os SNP normalmente são considerados **SNP comuns** em uma população se o alelo menos comum ocorre a uma frequência de aproximadamente 5% ou superior. Os SNP em relação aos quais o alelo menos comum ocorre a uma frequência inferior a 5% são considerados **SNP raros**. Em relação aos seres humanos, existe um SNP comum aproximadamente a cada 300 a

1.000 pb no genoma. É claro que existe um número muito maior de SNP raros.

Os SNP ocorrem dentro de genes, incluindo éxons, íntrons e regiões reguladoras. Os SNP em regiões codificadoras de proteínas podem ser classificados em um de três grupos: *sinônimos*, se os diferentes alelos codificam o mesmo aminoácido; *não sinônimos*, se os dois alelos codificam aminoácido diferentes; e *sem sentido*, se um alelo codifica um códon de fim e o outro, um aminoácido. Portanto, por vezes é possível associar um SNP a uma variação funcional nas proteínas e a uma alteração associada no fenótipo. Os SNP localizados fora de sequências codificadoras são denominados SNP *não codificadores* (ncSNP). Se os ncSNP não apresentam efeitos sobre a função gênica e o fenótipo, eles são denominados *silenciosos*. Os ncSNP silenciosos podem ser muito úteis na genética de populações, tendo em vista que podem ser utilizados como marcadores para abordar questões a respeito de processos da genética de populações, tais como o fluxo gênico entre populações.

Para estudar a variação de SNP em uma população, primeiramente precisamos determinar quais sítios de nucleotídios no genoma são variáveis — ou seja, constituem um SNP. Essa primeira etapa é denominada descoberta de SNP. Os SNP com frequência são descobertos por meio do sequenciamento de genomas de uma pequena amostra de indivíduos de uma espécie e em seguida com a comparação dessas sequências. Por exemplo, a descoberta de SNP em seres humanos teve início por meio do sequenciamento parcial de genomas de um **painel de descoberta** de 48 indivíduos de todo o mundo. Os sítios de nucleotídios variáveis foram descobertos por meio da comparação das sequências genômicas parciais desses 48 indivíduos entre si. Esse esforço inicial levou à descoberta de mais de 1 milhão de SNP.

Após a descoberta dos SNP, pode ser determinado o genótipo (a composição alélica) de diferentes indivíduos na população para cada SNP. Os microarranjos de DNA são uma tecnologia amplamente utilizada para essa finalidade (Figura 18.2). Os microarranjos utilizados para análises de SNP podem conter milhares de sondas correspondentes aos SNP conhecidos. Biotecnólogos desenvolveram diversos métodos diferentes para detectar variantes de SNP com a utilização de microarranjos. Em um método, o DNA de um indivíduo é marcado com corantes

fluorescentes e hibridizado com o microarranjo. Cada ponto (SNP) no microarranjo fluorescerá em vermelho para uma classe homozigota, em verde para a outra homozigota e em amarelo para o heterozigoto (ver Figura 18.2). O procedimento inteiro tem sido intensificado com a robótica para possibilitar a rápida *genotipagem* ou atribuição de genótipos (p. ex., A/A *versus* A/C) em grande escala.

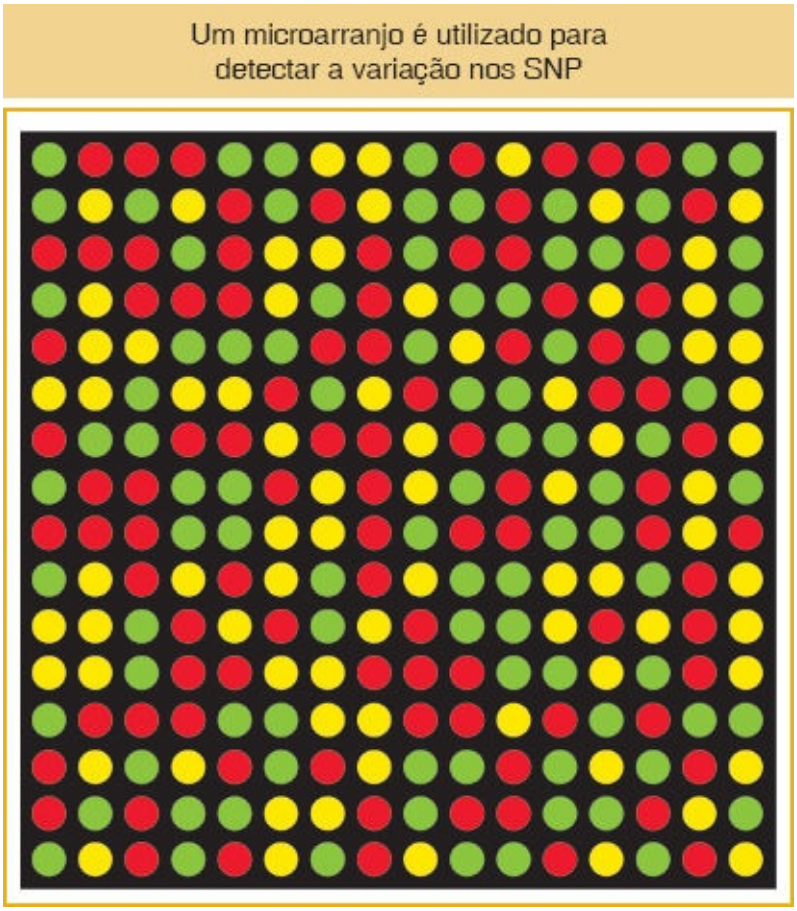


FIGURA 18.2 Detecção da variação no DNA: os SNP. Visualização de uma pequena parte de um microarranjo utilizado para realizar a varredura do genoma de um único indivíduo. Cada ponto representa um SNP, com *vermelho* e *verde* para as classes homozigotas e amarelo para os heterozigotos.

Microssatélites

Os microssatélites são *loci* poderosos para a análise da genética de populações por diversos motivos. Primeiramente, ao contrário dos SNP, que tipicamente apresentam apenas dois alelos por *locus* e que nunca podem apresentar mais de

quatro alelos, o número de alelos em um microssatélite com frequência é muito grande (20 ou mais). Em segundo lugar, eles apresentam uma alta taxa de mutação, tipicamente na faixa de 10^{-3} a 10^{-4} mutações por *locus* por geração, em comparação a 10^{-8} a 10^{-9} mutações por sítio por geração em relação aos SNP. A alta taxa de mutação significa que os níveis de variação são mais altos: mais alelos por *locus* e maior chance de que quaisquer dois indivíduos apresentem genótipos diferentes. Em terceiro lugar, os microssatélites são muito abundantes na maior parte dos genomas. Os seres humanos apresentam mais de um milhão de microssatélites.

Os microssatélites são observados por todo o genoma da maior parte dos organismos e podem estar presentes em éxons, íntrons, regiões reguladoras e sequências não funcionais de DNA. Os microssatélites com repetições de trinucleotídios são observados nas sequências codificadoras de alguns genes; esses codificam séries de um único aminoácido. O gene da doença de Huntington (*HD*) (ver Capítulo 16) contém uma repetição de CAG, que codifica uma série de glutaminas. Os indivíduos que carregam alelos com mais de 30 glutaminas são predispostos ao desenvolvimento da doença. Entretanto, em geral, a maior parte dos microssatélites está localizada fora das sequências codificadoras e a variação no número de repetições não está associada a diferenças no fenótipo.

São utilizados dois métodos principais para descobrir os *loci* de microssatélites no genoma. Se uma sequência genômica completa estiver disponível em relação a um organismo, pode-se simplesmente conduzir uma pesquisa para encontrá-los com a utilização de um computador. Em relação às espécies sem sequências genômicas (a maior parte dos organismos não modelo), é necessário um trabalho laboratorial considerável para descobrir microssatélites. Tipicamente, cria-se uma biblioteca genômica, realiza-se a varredura da biblioteca com uma sonda para o *motif* de interesse (p. ex., repetições AG) e determina-se a sequência de DNA dos clones selecionados para identificar os microssatélites e as sequências que os flanqueiam. Os métodos moleculares para a realização desse tipo de trabalho foram discutidos no Capítulo 10.

Após a identificação de um microssatélite e suas sequências flanqueadoras,

amostras de DNA de um conjunto de indivíduos na população podem ser analisadas para determinar o número de repetições que estão presentes em cada indivíduo. Para realizar a análise, são projetados *primers* de oligonucleotídios que correspondam às sequências flanqueadoras para a utilização na PCR. Se os *primers* forem marcados com um corante fluorescente, então os tamanhos dos produtos da PCR podem ser determinados no mesmo aparato utilizado para determinar a sequência de moléculas de DNA (Figura 18.3). Esses tamanhos revelam o número de repetições em um alelo de microssatélite. Por exemplo, o produto da PCR de um alelo de microssatélite que contém sete repetições AG será 8 pb mais longo do que um alelo que contém três repetições AG. Os indivíduos heterozigotos possuirão produtos de dois tamanhos diferentes. Tendo em vista que a PCR, a determinação do tamanho dos produtos da PCR e a classificação dos alelos podem ser, todas, automatizadas, é possível determinar os genótipos de grandes amostras de indivíduos em relação a grandes números de microssatélites de modo relativamente rápido.

Haplótipos

Para algumas questões na genética de populações, é importante considerar os genótipos de *loci* ligados como um grupo, em vez de individualmente. Os geneticistas utilizam o termo **haplótipo** para fazer referência à combinação de alelos em múltiplos *loci* no mesmo cromossomo homólogo. Dois cromossomos homólogos que compartilham o mesmo alelo em cada um dos *loci* em consideração têm o mesmo haplótipo. Se dois cromossomos apresentam genótipos diferentes, até mesmo em um dos *loci* em questão, eles então apresentam haplótipos diferentes. Se o *locus A* com alelos *A* e *a* estiver ligado ao *locus B* com alelos *B* e *b*, então existem quatro haplótipos possíveis em relação ao segmento cromossômico no qual esses *loci* estão localizados:

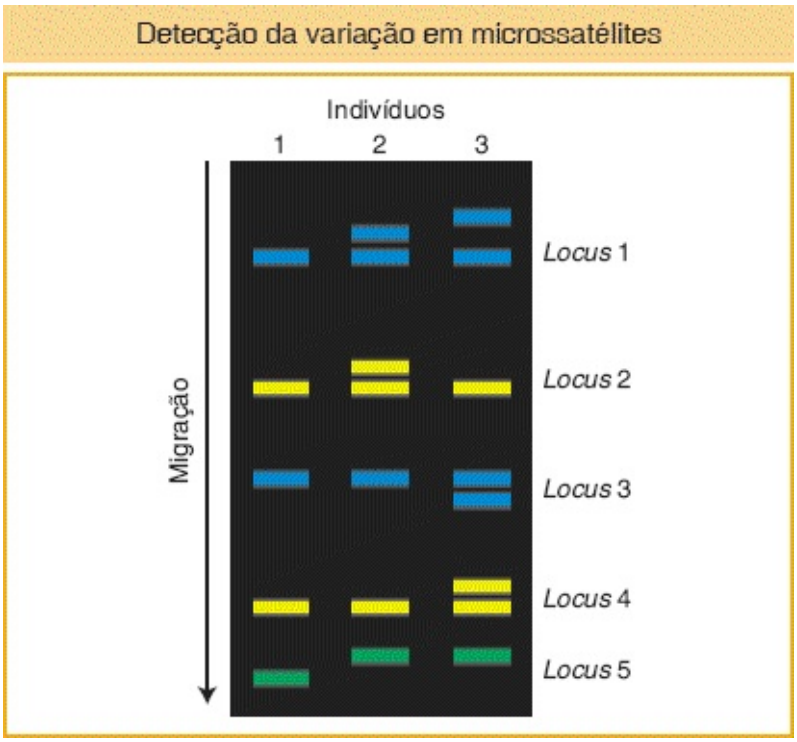


FIGURA 18.3 Detecção da variação no DNA: microssatélites. Desenho esquemático de uma imagem de gel dos *loci* em relação a cinco microssatélites avaliados simultaneamente. As três fileiras verticais correspondem a três indivíduos. Observe que existem três alelos presentes em relação ao *locus* 1 e que os indivíduos 2 e 3 são, ambos, heterozigotos em relação a este *locus*.

A	*B*
A	*b*
a	*B*
a	*b*

Um exemplo mais complexo, porém mais realista, está demonstrado na Figura 18.4. Na Figura 18.4 A, existem sete segmentos cromossômicos, mas apenas seis haplótipos, tendo em vista que os segmentos cromossômicos 5 e 6 apresentam o mesmo haplótipo (E).

Os haplótipos são utilizados com mais frequência na genética de populações em relação aos *loci* que estão fisicamente próximos. Por exemplo, os sítios de nucleotídios variáveis em um único gene podem ser utilizados para definir

haplótipos em relação àquele gene. Entretanto, o conceito de haplótipo funciona em relação a regiões maiores quando há pouca ou nenhuma recombinação ao longo da região. Ele pode até mesmo ser aplicado para um cromossomo inteiro, tal como o cromossomo Y humano. Finalmente, por vezes ele é útil para agrupar haplótipos em classes. Conforme demonstrado na Figura 18.4 A, existem duas classes principais de haplótipos (I e II), que diferem em cinco sítios de nucleotídios mais um microssatélite. Entretanto, cada classe contém diversos subtipos (I-a, I-b, …). A **rede de haplótipos** demonstra as relações entre os haplótipos, posicionando cada mutação em um dos ramos (Figura 18.4 B).

Quais percepções podemos obter a partir da análise de haplótipos? Os geneticistas de populações estudando o cromossomo Y humano entre homens asiáticos descobriram um haplótipo altamente prevalente, denominado haplótipo "aglomerado estelar" (Figura 18.5 A). Tipicamente, a maior parte dos homens apresenta um haplótipo de cromossomo Y raro, mas o haplótipo "aglomerado estelar" está presente em 8% dos homens asiáticos. Com a utilização da taxa de mutação conhecida, os pesquisadores estimaram que esse haplótipo comum surgiu entre 700 e 1.300 anos atrás. (Posteriormente neste capítulo, discutiremos as taxas de mutação e sua utilização na genética de populações.) Esse haplótipo é mais comum na Mongólia, sugerindo que ele surgiu ali. Os pesquisadores inferiram que o haplótipo "aglomerado estelar" remonta a um homem na Mongólia há aproximadamente 1.000 anos. Notavelmente, a distribuição atual desse haplótipo segue os limites geográficos do Império Mongol estabelecido por Genghis Khan há aproximadamente 1.200 anos (Figura 18.5 B). Aparentemente todos os homens contemporâneos com esse haplótipo são descendentes de Genghis Khan (ou de seus parentes da linhagem masculina).

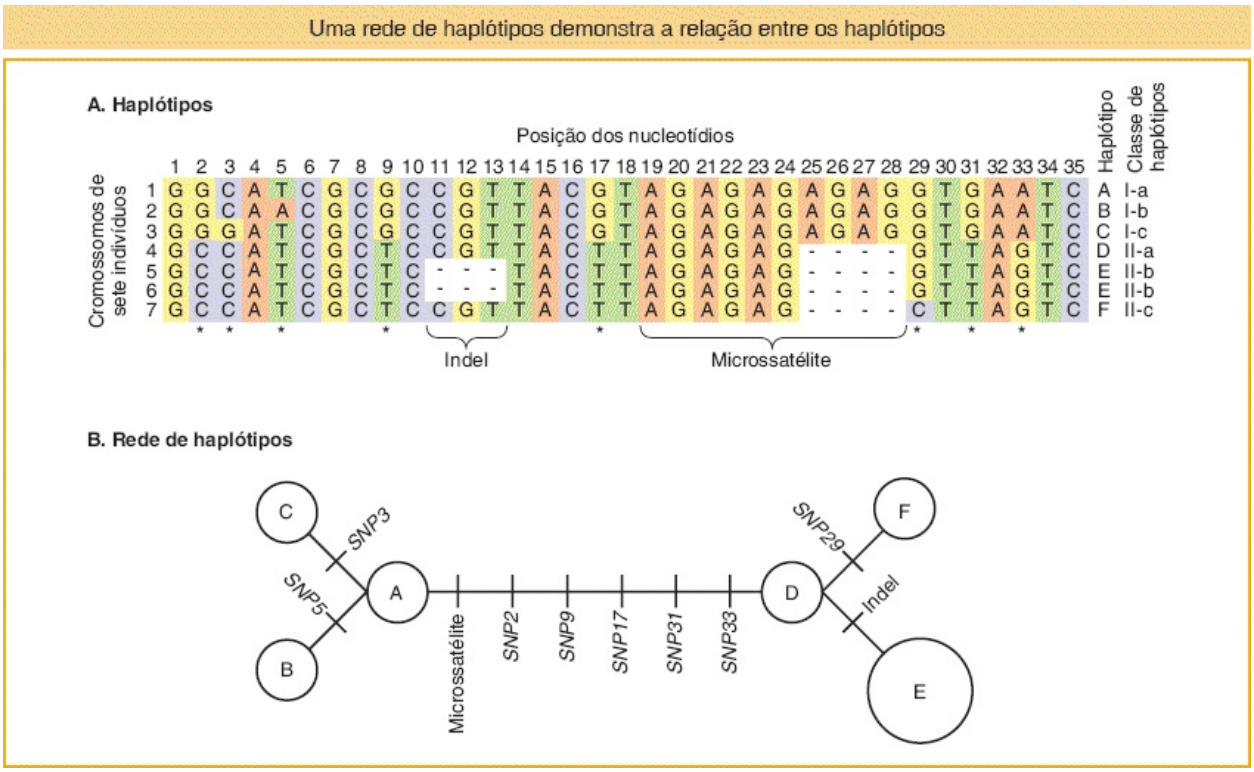


FIGURA 18.4 A. Existe um total de seis haplótipos (A a F) nas sequências de DNA alinhadas de sete cromossomos individuais de diferentes pessoas. B. Estes seis haplótipos estão unidos em uma rede de haplótipos que demonstra as relações entre os haplótipos. Cada círculo representa um dos seis haplótipos. Quaisquer dois haplótipos diferem nos *loci* observados em todos os ramos que os conectam. Os asteriscos demonstram a localização dos SNP.

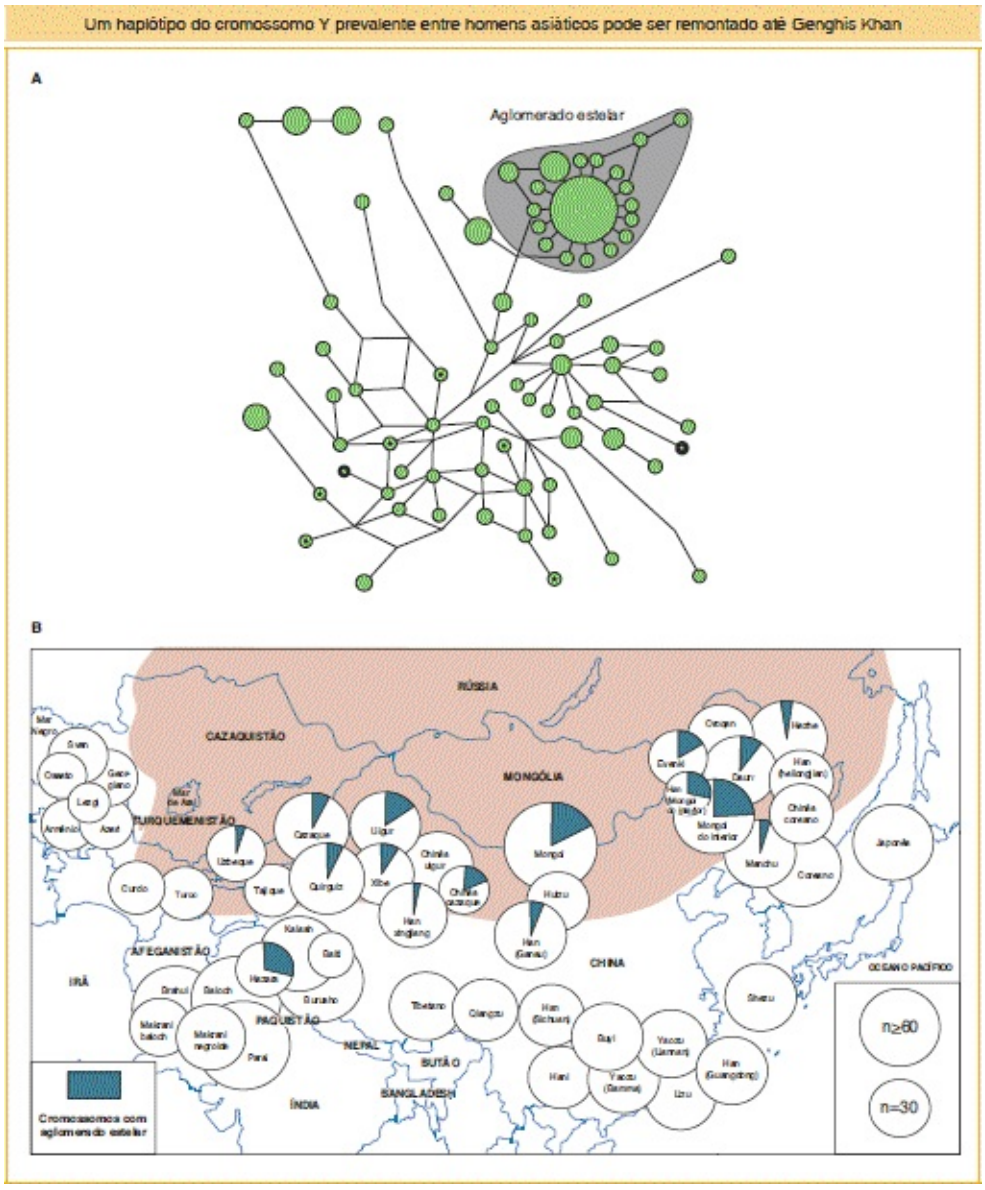


FIGURA 18.5 A. Rede de haplótipos em relação aos cromossomos Y de homens asiáticos, demonstrando a predominância do haplótipo aglomerado estelar que se acredita remontar a Genghis Khan. A área do *círculo* é proporcional ao número de indivíduos com o haplótipo específico que o círculo representa. **B.** Distribuição geográfica do haplótipo aglomerado estelar. As populações estão demonstradas como *círculos*, com uma área proporcional ao tamanho da amostra; a proporção de indivíduos na amostra que carregam cromossomos com aglomerado estelar está indicada pelos *setores verdes*. Não foram observados cromossomos com aglomerado estelar nas populações que não apresentam *setor verde no círculo*. A *área sombreada* representa a extensão do império de Genghis Khan. (Dados de T. Zerjal *et al.*, *Am. J. Hum. Genet.* 72, 2003, 717-721.)

Outras fontes e outros tipos de variação

Além dos SNP e dos microssatélites, qualquer variação na sequência de DNA dos

cromossomos em uma população é passível de análise genética de populações. As variações que podem ser analisadas incluem inversões, translocações, deleções ou duplicações, e a presença ou a ausência de um elemento de transposição em um *locus* em particular no genoma. Outro tipo comum de variação é o polimorfismo por inserção e deleção ou *indel*, abreviadamente (ver Capítulo 16). Esse tipo de polimorfismo envolve a presença ou a ausência de um ou mais nucleotídios em um *locus* em um alelo em relação a outro. Na Figura 18.1, os segmentos cromossômicos 5 e 6 diferem dos outros cinco segmentos por uma indel de 3 pb. Contrariamente aos microssatélites, as indels não contêm *motifs* repetidos, tais como AGAGAGAG.

Até agora, a nossa discussão sobre os SNP e os microssatélites enfocou o genoma nuclear. Entretanto, também pode ser observada uma interessante variação genética nos genomas mitocondrial (mtDNA) e de cloroplasto (cpDNA) de eucariotos. Tanto os SNP quanto os microssatélites são observados nesses genomas de organelas. Tendo em vista que o mtDNA e o cpDNA normalmente são herdados por via materna, sua análise pode ser utilizada para seguir a história das linhagens femininas. Em 1987, um proeminente estudo da linhagem mitocondrial humana traçou a história dos haplótipos do mtDNA humano e determinou que os genomas mitocondriais de todos os seres humanos modernos remontam a uma única mulher que viveu na África há aproximadamente 150.000 anos (Figura 18.6). Ela foi apelidada de “Eva Mitocondrial” pela imprensa popular. Esse estudo do mtDNA foi a primeira análise genética completa a sugerir que todos os seres humanos modernos vieram da África.

Projeto HapMap

Um importante avanço na genética de populações humanas ao longo da última década foi a criação de um mapa de haplótipos de todo o genoma, ou **HapMap**. Um consórcio de cientistas ao redor do mundo genotipou milhares de pessoas que representam a diversidade de nossa espécie em relação a centenas de milhares de SNP e microssatélites. O resultado é um quadro altamente detalhado da variação em nossa espécie. Os dados estão disponíveis para o público em diversos *web sites*, incluindo aquele do International HapMap Project (www.hapmap.org) e do

Human Genome Diversity Project (hgdp.uchicago.edu). Neste capítulo, utilizaremos esses dados para apresentar os princípios da genética de populações. Embora primeiramente desenvolvidos para seres humanos, os HapMaps têm sido, desde então, desenvolvidos para diversas outras espécies, incluindo *Drosophila*, camundongo, *Arabidopsis*, arroz e milho.

> CONCEITO-CHAVE Os genomas estão repletos de diversos tipos de variação passíveis de análise por genética de populações. Os SNP e os microssatélites são os dois tipos de polimorfismos mais comumente estudados na genética de populações. Tecnologias de alta produção possibilitam que centenas de milhares de polimorfismos sejam classificados em dezenas de milhares de indivíduos.

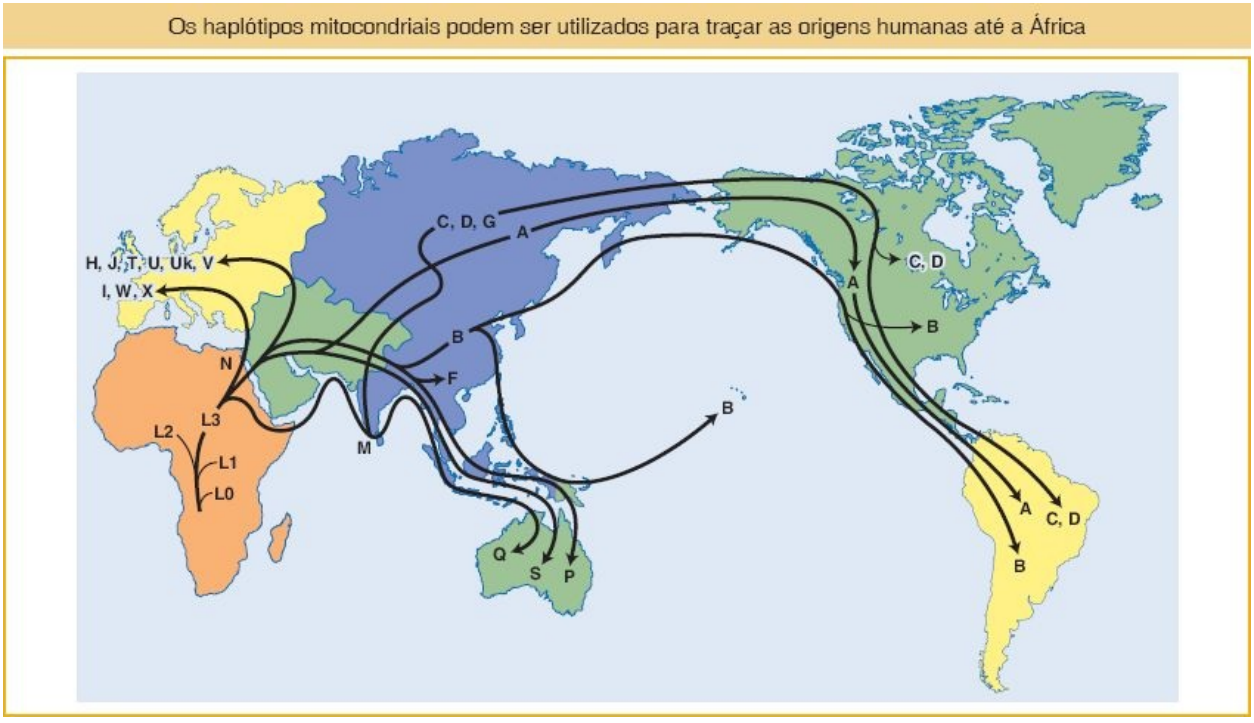


FIGURA 18.6 A rede de haplótipos para grupos de haplótipos do mtDNA desenhada em um mapa mundial. O grupo de haplótipo L ancestral aparece na África e os grupos derivados (A, B e assim por diante) estão dispersos por todo o mundo. (Dados de www.mitomap.org.)

18.2 Conceito de pool gênico e lei de Hardy-Weinberg

Talvez você tenha visto alguém realizando um truque mortal e pensado que a pessoa sofria o risco de eliminar a si própria do "*pool* gênico". Em caso afirmativo, você estava utilizando um conceito, o *pool* gênico, que tem sua origem diretamente na genética de populações e que trilhou o seu caminho até a cultura popular. O conceito de *pool* gênico é uma ferramenta básica para se pensar a respeito da variação genética nas populações. Podemos definir o ***pool* gênico** como a soma total de todos os alelos nos membros reprodutivos de uma população em um determinado momento. Por exemplo, a Figura 18.7 demonstra uma população de 16 rãs, cada uma das quais carregando dois alelos no *locus* autossômico *A*. Por meio da simples contagem, podemos determinar que existem cinco homozigotos *A/A*, oito heterozigotos *A/a* e três homozigotos *a/a*. O tamanho da população, normalmente simbolizado pela letra *N*, é 16, e existe um total de 32 alelos, ou 2*N*, nessa população diploide. Com esse simples conjunto de números, descrevemos o *pool* gênico em relação ao *locus A*.

Tipicamente, os geneticistas de populações não se importam com contagens absolutas dos diferentes genótipos em uma população, mas sim com as **frequências genotípicas**. Podemos calcular a frequência do genótipo *A/A* simplesmente dividindo o número de indivíduos *A/A* pelo número total de indivíduos na população (*N*), para obter 0,31. A frequência de heterozigotos *A/a* é de 0,50, e a frequência de homozigotos *a/a* é de 0,19. Por serem frequências, elas somam 1,0. As frequências são uma medida mais prática do que as contagens absolutas, tendo em vista que raramente os geneticistas de populações são capazes de estudar todos os indivíduos em uma população. Em vez disso, os geneticistas de populações coletarão uma amostra aleatória ou não tendenciosa de indivíduos em uma população e utilizarão a amostra para inferir as frequências genotípicas na população inteira.

FIGURA 18.7 Um *pool* gênico de rãs.

Podemos fazer uma descrição mais simples desse *pool* gênico de rãs se calcularmos as **frequências alélicas** em vez das frequências genotípicas (Quadro 18.1). Na Figura 18.7, 18 dos 32 alelos são *A* e, assim, a frequência de *A* é 18/32 = 0,56. A frequência do alelo *A* é tipicamente simbolizada pela letra *p* e, nesse caso, *p* = 0,56. A frequência do alelo *a* é simbolizada pela letra *q* e, nesse caso, *q* = 14/32 = 0,44. Novamente, tendo em vista que essas são frequências, a sua soma é 1,0: *p* + *q* = 0,56 + 0,44 = 1,0. Agora temos uma descrição do nosso *pool* gênico de rãs com a utilização de apenas dois números, *p* e *q*.

> CONCEITO-CHAVE O *pool* gênico é um conceito fundamental para o estudo da variação genética nas populações: ele é a soma de todos os alelos nos membros reprodutivos de uma população em uma determinada ocasião. Podemos descrever a variação em uma população em termos das frequências genotípicas e alélicas.

Conforme mencionado anteriormente, um objetivo importante da genética de populações é compreender a transmissão de alelos de uma geração até a próxima nas populações naturais. Nesta seção, começaremos a ver como isso funciona. Veremos como podemos utilizar as frequências alélicas no *pool* gênico para fazer previsões a respeito das frequências genotípicas na próxima geração.

Quadro 18.1 Cálculo das frequências alélicas.

Em um *locus* com dois alelos *A* e *a*, definiremos as frequências dos três genótipos *A/A*, *A/a* e *a/a* como $f_{A/A}$, $f_{A/a}$ e $f_{a/a}$, respectivamente. Podemos utilizar essas frequências genotípicas para calcular as frequências alélicas: *p* é a frequência do alelo *A* e *q* é a frequência do alelo *a*. Tendo em vista que cada homozigoto *A/A* é composto apenas por alelos *A* e tendo em vista que metade dos alelos de cada heterozigoto *A/a* é de alelos *A*, a frequência total *p* de alelos *A* na população é calculada como:

$$p = f_{A/A} + \tfrac{1}{2} f_{A/a} = \text{Frequência de A}$$

De modo semelhante, a frequência *q* do alelo *a* é fornecida por:

$$q = f_{a/a} + \tfrac{1}{2} f_{A/a} = \text{Frequência de } a$$

Portanto:

$$p + q = f_{A/a} + f_{A/a} + f_{a/a} = 1{,}0$$

e:

$$q = 1 - p$$

Se houver mais de duas formas alélicas diferentes, a frequência em relação a cada alelo é simplesmente a frequência de seu homozigoto mais metade da soma das frequências de todos os heterozigotos nos quais ele aparece.

A frequência de um alelo no *pool* gênico é igual à probabilidade de que o alelo seja escolhido na coleta aleatória de um alelo do *pool* gênico para formar um ovócito ou um espermatozoide. Sabendo disso, podemos calcular a probabilidade de que uma rã na próxima geração seja um homozigoto *A/A*. Se nos dirigirmos ao *pool* gênico das rãs (ver Figura 18.7) e coletarmos o primeiro alelo, a probabilidade de que ele seja um *A* é $p = 0{,}56$ e, de modo semelhante, a probabilidade de que o segundo alelo que coletarmos também seja um *A* é $p = 0{,}56$. O produto dessas duas probabilidades, ou $p^2 = 0{,}3136$, é a probabilidade de que uma rã na próxima geração seja *A/A*. A probabilidade de que uma rã na próxima geração seja *a/a* é $q^2 = 0{,}44 \times 0{,}44 = 0{,}1936$. Existem dois modos para produzir um heterozigoto. Podemos primeiramente coletar um *A* com probabilidade *p* e em seguida coletar um *a* com probabilidade *q*, ou podemos coletar o *a* primeiramente e o *A* em segundo lugar. Portanto, a probabilidade de que uma rã na próxima geração seja heterozigota *A/a* é $pq + qp = 2pq = 0{,}4928$. Em geral, as frequências (*f*) genotípicas são:

$f_{A/A} = p^2$
$f_{a/a} = q^2$
$f_{A/a} = 2pq$

Finalmente, conforme esperado, a soma da probabilidade de ser *A/A*, mais a probabilidade de ser *A/a*, mais a probabilidade de ser *a/a* é 1,0:

$$p^2 + 2pq + q^2 = 1{,}0$$

Essa equação simples é a **lei de Hardy-Weinberg**, um dos fundamentos da teoria da genética de populações.

O processo de acesso ao *pool* gênico para coletar um alelo é denominado *amostragem* do *pool* gênico. Tendo em vista que qualquer indivíduo que contribui para o *pool* gênico pode produzir muitos ovócitos ou espermatozoides que carreiem exatamente a mesma cópia de um alelo, é possível coletar uma cópia em particular e em seguida voltar ao *pool* gênico e coletar exatamente a mesma cópia mais uma vez. Existe também um elemento de acaso envolvido na amostragem do

pool gênico. Algumas cópias podem, por acaso, ser coletadas mais de uma vez e outras podem absolutamente não ser coletadas. Posteriormente no capítulo, veremos como essas propriedades da amostragem do *pool* gênico podem levar a alterações no *pool* gênico ao longo do tempo.

Utilizamos a lei de Hardy-Weinberg para calcular as frequências genotípicas na próxima geração a partir das frequências alélicas na geração atual. Também podemos utilizar a lei de Hardy-Weinberg para calcular as frequências alélicas a partir das frequências genotípicas em uma única geração. Por exemplo, alguns tipos de albinismo em seres humanos ocorrem em virtude de alelos recessivos no *locus* *OCA2*. Na África, um tipo de albinismo, denominado albinismo oculocutâneo marrom, resulta de um alelo recessivo de *OCA2* (Figura 18.8). Os indivíduos com essa condição estão presentes a frequências tão altas quanto 1 em 1.100 em alguns grupos étnicos na África. Podemos utilizar a lei de Hardy-Weinberg para calcular as frequências alélicas:

$$f_{a/a} = q^2 = 1/1.100 = 0,0009$$

assim:
e:

$$q = \sqrt{0,0009} = 0,03$$

$$p = 1 — q = 0,97$$

Com a utilização das frequências alélicas, também podemos calcular a frequência de heterozigotos na população como:

$$2pq = 2 \times 0,97 \times 0,03 = 0,06$$

O último número prevê que aproximadamente 6% dessa população são heterozigotos ou portadores do alelo recessivo no *OCA2*.

Quando utilizamos a lei de Hardy-Weinberg para calcular as frequências alélicas ou genotípicas, realizamos algumas presunções críticas.

- *Primeiramente*, presumimos que o cruzamento é aleatório na população quanto ao gene em questão. O desvio do cruzamento aleatório viola essa presunção, tornando inadequada a aplicação da lei de Hardy-Weinberg. Por exemplo, uma tendência de indivíduos fenotipicamente semelhantes de cruzarem entre si viola a lei de Hardy-Weinberg. Se albinos cruzassem mais frequentemente com outros albinos do que com não albinos, então a lei de Hardy-Weinberg superestimaria a frequência do alelo recessivo

FIGURA 18.8 Indivíduo de ancestralidade africana com albinismo oculocutâneo marrom (BOCA), uma condição definida pela pele bronze-clara e pelos cabelos bege a marrom-claros. (Dra. Michele Ramsay, Department of Human Genetics, School of Pathology, the National Health Laboratory Service University of Witwatersrand.)

- *Em segundo lugar*, se um dos genótipos reduzir a viabilidade de tal modo que alguns indivíduos com aquele genótipo morrem antes que as frequências dos genótipos sejam contadas, então a estimativa das frequências gênicas será imprecisa
- *Em terceiro lugar*, para que a lei de Hardy-Weinberg seja aplicada, a

população não pode ser dividida em subpopulações que sejam parcial ou totalmente isoladas geneticamente. Se existirem subpopulações separadas, os alelos podem estar presentes em frequências distintas nas diferentes subpopulações. Nesse caso, a utilização das contagens genotípicas da população em geral pode não fornecer uma estimativa precisa das frequências alélicas globais

- *Finalmente*, a lei de Hardy-Weinberg se aplica estritamente apenas às populações infinitamente grandes. Em relação às populações finitas, haverá desvios das frequências previstas pela lei de Hardy-Weinberg em virtude do acaso quando a amostragem do *pool* gênico produzir a próxima geração.

Observamos como podemos utilizar a lei de Hardy-Weinberg e as frequências gênicas na geração atual (t_0) para calcular as frequências genotípicas na próxima geração (t_1) por meio da amostragem aleatória do *pool* gênico para a produção de ovócitos e espermatozoides. De modo semelhante, as frequências genotípicas previstas em relação à geração t_1 podem ser utilizadas, por sua vez, para calcular as frequências gênicas em relação à próxima geração (t_2). As frequências gênicas na geração t_2 permanecerão as mesmas da geração t_1. Sob a lei de Hardy-Weinberg, nem as frequências gênicas nem as genotípicas mudam de uma geração para a próxima quando uma população infinitamente grande é amostrada aleatoriamente para a formação de ovócitos e espermatozoides. Portanto, uma lição importante da lei de Hardy-Weinberg é que, em populações grandes, a variação genética não é criada nem destruída pelo processo de transmissão de genes de uma geração para a próxima. Diz-se que as populações que aderem a esse princípio estão em **equilíbrio de Hardy-Weinberg**.

	Frequências genotípicas			Frequências gênicas	
Geração	*A/A*	*A/a*	*a/a*	*A*	*a*
t_0	0,64	0,32	0,04	0,8	0,2

t_1	0,64	0,32	0,04	0,8	0,2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
t_n	0,64	0,32	0,04	0,8	0,2

Aqui estão mais alguns poucos pontos a respeito da lei de Hardy-Weinberg.

1. Em relação a qualquer alelo que exista em uma frequência muito baixa, os indivíduos homozigotos serão observados apenas muito raramente. Se o alelo *a* apresenta uma frequência de 1 em mil ($q = 0,001$), então apenas 1 em 1 milhão (q^2) de indivíduos será homozigoto em relação àquele alelo. Consequentemente, os alelos recessivos para distúrbios genéticos podem ocorrer no estado heterozigoto em muito mais indivíduos do que aqueles que de fato expressam o distúrbio genético em questão.

2. A lei de Hardy-Weinberg ainda é aplicável quando existem mais de dois alelos por *locus*. Se existem *n* alelos, A_1, A_2,… A_n, com as frequências p_1, p_2, … p_n, então a soma de todas as frequências individuais é igual a 1,0. As frequências de cada um dos genótipos homozigotos são simplesmente o quadrado das frequências dos alelos, e as frequências das diferentes classes de heterozigotos são duas vezes o produto das frequências do primeiro e do segundo alelo. A Tabela 18.1 fornece um exemplo com $p_1 = 0,5$, $p_2 = 0,3$ e $p_3 = 0,2$.

3. A lógica de Hardy-Weinberg também é aplicável aos *loci* ligados ao X. Os homens são hemizigotos em relação aos genes ligados ao X, o que significa que um homem apresenta uma única cópia desses genes. Portanto, em relação

aos genes ligados ao X em homens, as frequências genotípicas são iguais às frequências alélicas. Para as mulheres, as frequências genotípicas em relação aos genes ligados ao X seguem as expectativas normais de Hardy-Weinberg.

A calvície masculina padrão é um traço ligado ao X (Figura 18.9). O gene *AR* (em referência a receptor de andrógeno) é um gene ligado ao X envolvido no desenvolvimento masculino. Existe um haplótipo *AR* denominado *Eur-H1* que está fortemente associado à calvície-padrão. A calvície masculina padrão é comum na Europa, onde o haplótipo *Eur-H1* ocorre a uma frequência de 0,71, o que significa que 71% dos homens europeus o carregam. Com a utilização da lei de Hardy-Weinberg, podemos calcular que 50% das mulheres europeias são homozigotas para o *Eur-H1* e que 41% são heterozigotas. A herança da calvície é complexa e é afetada por múltiplos genes. Assim, nem todos os homens que apresentam o *Eur-H1* se tornam calvos.

4. Pode-se testar se as frequências genotípicas observadas em um *locus* correspondem às previsões de Hardy-Weinberg com a utilização do teste de χ^2 (ver Capítulo 3). Um exemplo é fornecido pelo gene do antígeno leucocitário humano, *HLA-DQA1*, do complexo principal de histocompatibilidade (MHC). O MHC é um agrupamento de genes no cromossomo 6 que desempenham papéis no sistema imune. A Tabela 18.2 apresenta as frequências genotípicas em relação a um SNP (rs9272426) no *HLA-DQA1* para 84 habitantes da Toscana, na Itália. Esse SNP apresenta os alelos *A* e *G*. A partir das frequências genotípicas na Tabela 18.2, podemos calcular as frequências alélicas: $f(\text{A}) = p = 0{,}53$ e $f(\text{G}) = q = 0{,}47$. Em seguida, podemos calcular as frequências genotípicas esperadas sob a lei de Hardy-Weinberg: $p^2 = 0{,}281$, $2pq = 0{,}498$ e $q^2 = 0{,}221$. A multiplicação das frequências genotípicas esperadas pelo tamanho da amostra ($N = 84$) nos fornece o número esperado de indivíduos para cada genótipo. Agora, podemos calcular a estatística do χ^2 como 8,29. Com a utilização da Tabela 3.1, observamos que a probabilidade sob a hipótese nula de que os dados observados correspondam às previsões de Hardy-Weinberg é $P < 0{,}005$,

com gl = 1. (Temos apenas um grau de liberdade, tendo em vista que temos três categorias genotípicas e utilizamos dois números a partir dos dados [*N* e *p*] para calcular os valores esperados [3 — 2 resulta em 1 grau de liberdade]. Não precisamos utilizar *q*, tendo em vista que *q* = *p* — 1.) Essa análise nos conduz à forte suspeita de que os toscanos não estão em conformidade com as expectativas de Hardy-Weinberg em relação ao *HLA-DQA1*. Observaremos adicionalmente a genética de populações do MHC na Seção 18.3 sobre os sistemas de acasalamento e na Seção 18.5 sobre a seleção natural.

Tabela 18.1 Frequências genotípicas de Hardy-Weinberg em relação a um *locus* com três alelos A_1, A_2, e A_3 com frequências 0,5, 0,3 e 0,2, respectivamente.

Genótipo	Expectativa	Frequência
A_1A_1	p_1^2	0,25
A_2A_2	p_2^2	0,09
A_3A_3	p_3^2	0,04
A_1A_2	$2\ p_1\ p_2$	0,30
A_1A_3	$2\ p_1\ p_3$	0,20
A_2A_3	$2\ p_2\ p_3$	0,12
Soma		1,00

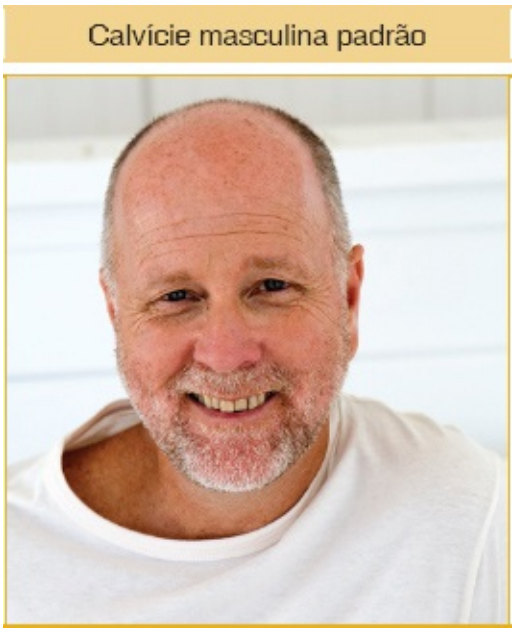


FIGURA 18.9 Indivíduo demonstrando calvície masculina padrão, uma condição ligada ao cromossomo X. (B2M Productions/Getty Images.)

Tabela 18.2 Frequências de genótipos do SNP rs9272426 no *locus HLA-DQA1* do MHC para pessoas da Toscana, na Itália.

	Genótipos			
	A/A	**A/G**	**G/G**	**Soma**
Número observado	17	55	12	84
Frequência observada	0,202	0,655	0,143	1
Frequência esperada	0,281	0,498	0,221	1
Número observado	23,574	41,851	18,574	84

(Observado — esperado)²/esperado	1,833	4,131	2,327	8,29

Fonte: International HapMap Project (www.hapmap.org).

A lei de Hardy-Weinberg é parte do fundamento da genética de populações. Ela se aplica a uma população idealizada de tamanho infinito e na qual o cruzamento é aleatório. Ela também presume que todos os genótipos são igualmente adaptados — ou seja, que todos eles são igualmente viáveis e que apresentam o mesmo sucesso na reprodução. As populações reais se desviam dessa idealizada. No restante do capítulo, examinaremos como fatores tais como o cruzamento não aleatório, o tamanho de população finito e a aptidão desigual de diferentes genótipos causam desvios das expectativas de Hardy-Weinberg. Também veremos como a lei de Hardy-Weinberg pode ser modificada para compensar esses fatores.

> CONCEITO-CHAVE A lei de Hardy-Weinberg descreve a relação entre as frequências alélicas e genotípicas. Essa lei nos informa que a variação genética não é criada nem destruída pelo processo de transmissão de genes de uma geração para a próxima. A lei de Hardy-Weinberg se aplica estritamente a populações infinitamente grandes e de cruzamento aleatório.

18.3 Sistemas de acasalamento

O cruzamento aleatório é uma suposição crítica da lei de Hardy-Weinberg. A suposição de cruzamento aleatório é atendida se todos os indivíduos na população apresentam igual probabilidade de representar uma escolha quando um parceiro é escolhido. Entretanto, se um parente, vizinho ou indivíduo fenotipicamente semelhante é um parceiro mais provável do que um indivíduo aleatório, então a suposição de cruzamento aleatório foi violada. As populações que não têm cruzamento aleatório não exibirão as proporções de Hardy-Weinberg

exatas em relação aos genótipos em alguns dos genes, ou todos. Três tipos de desvios na escolha do parceiro que violam a suposição de cruzamento aleatório são o cruzamento preferencial, o isolamento pela distância e o endocruzamento.

Cruzamento preferencial

Ocorre o cruzamento preferencial se os indivíduos escolhem parceiros com base na semelhança consigo mesmos. O **cruzamento preferencial positivo** ocorre quando tipos semelhantes acasalam; por exemplo, se indivíduos altos acasalam preferencialmente com outros indivíduos altos e indivíduos baixos acasalam com outros indivíduos baixos. Nesses casos, os genes que controlam a diferença na altura não seguirão a lei de Hardy-Weinberg. Em vez disso, esperaríamos observar um excesso de homozigotos em relação aos alelos "altos" entre a progênie de pares de cruzamento altos e um excesso de homozigotos em relação aos alelos "baixos" entre a progênie de pares de cruzamento baixos. Em seres humanos, ocorre um cruzamento preferencial positivo em relação à altura.

O **cruzamento preferencial negativo** ou **não preferencial** ocorre quando indivíduos diferentes acasalam — ou seja, quando os opostos se atraem. Um exemplo de cruzamento preferencial negativo é fornecido pelo *locus* da autoincompatibilidade, ou *S*, em plantas como *Brassica* (brócolis e seus parentes). Existem diversos alelos no *locus S*, S_1, S_2, S_3 e assim por diante. O estigma de uma planta não será receptivo ao pólen que carrega qualquer um de seus dois próprios alelos (Figura 18.10). Por exemplo, o estigma de um heterozigoto S_1/S_2 não possibilitará que os grãos de pólen que carregam um alelo S_1 ou S_2 germinem e fertilizem os seus óvulos, embora os grãos de pólen que carregam os alelos S_3 ou S_4 possam fazer isso. Esse mecanismo bloqueia a autofertilização, forçando, assim, a polinização cruzada. O *locus S* viola a lei de Hardy-Weinberg, uma vez que não são formados genótipos homozigotos para *S*.

Um segundo exemplo de cruzamento preferencial negativo é fornecido pelo complexo principal de histocompatibilidade (MHC), que sabidamente influencia a escolha do parceiro em vertebrados. O MHC afeta o odor corporal em camundongos e ratos, proporcionando uma base para a escolha do parceiro. Nos que são conhecidos como os "experimentos da camiseta suada", pesquisadores

solicitaram a um grupo de homens que vestissem camisetas durante 2 dias. Em seguida, eles solicitaram a um grupo de mulheres que sentissem o odor das camisetas e as classificassem em relação à "preferência". As mulheres preferiram o odor de homens cujos haplótipos do MHC eram diferentes dos seus próprios. Os dados do projeto HapMap humano confirmaram que os casais americanos são significativamente mais heterozigotos quanto ao MHC do que o esperado ao acaso. O MHC desempenha um papel central na nossa resposta imune aos patógenos, e os heterozigotos podem ser mais resistentes aos patógenos. Portanto, a nossa descendência se beneficia se acasalamos de modo não preferencial em relação ao nosso genótipo MHC. Esse mecanismo pode explicar o motivo de o SNP no gene *HLA-DQA1* de MHC a respeito do qual discutimos anteriormente não seguir a lei de Hardy-Weinberg entre os residentes da Toscana. Observe novamente a Tabela 18.2 e você notará que existem mais heterozigotos do que o esperado, 55 *versus* 42. Os toscanos aparentam estar praticando o cruzamento não preferencial em relação a esse SNP.

FIGURA 18.10 Cruzamento preferencial negativo causado pelo *locus* de autoincompatibilidade (*S*) do gênero de plantas florescentes *Brassica*. **A.** Um estigma autopolinizado S_1/S_2 demonstra ausência de crescimento do tubo de pólen. **B.** Ocorre o crescimento do tubo de pólen para um estigma S_1/S_2 submetido à polinização cruzada com o pólen de um heterozigoto S_3/S_4. (June Bowman Nasrallah.)

Isolamento pela distância

Outro tipo de desvio na escolha do parceiro tem origem a partir da distância geográfica entre os indivíduos. Os indivíduos são mais aptos a acasalarem com um vizinho do que com outro membro da sua espécie do lado oposto do continente — ou seja, os indivíduos podem demonstrar **isolamento pela distância**. Consequentemente, as frequências alélicas e genotípicas com frequência são diferentes entre peixes em lagos separados ou entre pinheiros em diferentes

regiões de um continente. Diz-se que as espécies ou as populações que exibem tal padrão de variação genética demonstram **estrutura populacional**. Uma espécie pode ser dividida em uma série de subpopulações, tais como as rãs em diferentes lagoas, ou as pessoas em diferentes cidades.

Se uma espécie apresenta estrutura populacional, a proporção de homozigotos será maior na espécie como um todo do que o esperado sob a lei de Hardy-Weinberg. Considere um exemplo hipotético de uma espécie de girassóis selvagens distribuída por todo o estado do Kansas com um gradiente na frequência do alelo *A* de 0,9 perto de Kansas City a 0,1 perto de Elkhart (Figura 18.11 A). Obtivemos amostras de 100 plantas de girassol de cada uma dessas duas cidades, mais 100 de Hutchinson, no centro do estado, e calculamos as frequências alélicas. Cada cidade representa uma subpopulação. Em relação a qualquer uma das três cidades, a lei de Hardy-Weinberg funciona bem. Por exemplo, em Elkhart, esperamos que $Nq^2 = 100 \times (0,9)^2 = 81$ homozigotos *a/a*, e isso é o que observamos. Entretanto, para todo o estado, preveríamos que $Nq^2 = 300 \times (0,5)^2 = 75$ homozigotos *a/a*, mas observamos 107. Em virtude da estrutura populacional, existem mais plantas de girassol homozigotas do que o esperado.

	Número de indivíduos					
	N	*A/A*	*A/a*	*a/a*	*p*	*q*
Kansas City	100	81	18	1	0,90	0,10
Hutchinson	100	25	50	25	0,50	0,50
Elkhart	100	1	18	81	0,10	0,90
Todo o						

estado (observado)	300	107	86	107	0,50	0,50
Todo o estado (esperado)	300	75	150	75	—	—

Eis um exemplo real de estrutura populacional de nossa própria espécie. Na África, o alelo FY^{nulo} do grupo sanguíneo Duffy demonstra um gradiente com uma baixa frequência no leste e no norte da África uma frequência moderada no sul da África, e uma alta frequência ao longo da África central (Figura 18.11 B). Esse alelo é raro fora da África. Em virtude desse gradiente, não podemos utilizar as frequências alélicas globais na África para calcular as frequências genotípicas com a utilização da lei de Hardy-Weinberg. Posteriormente neste capítulo e no Capítulo 20, discutiremos a relação entre o FY^{nulo} e a malária.

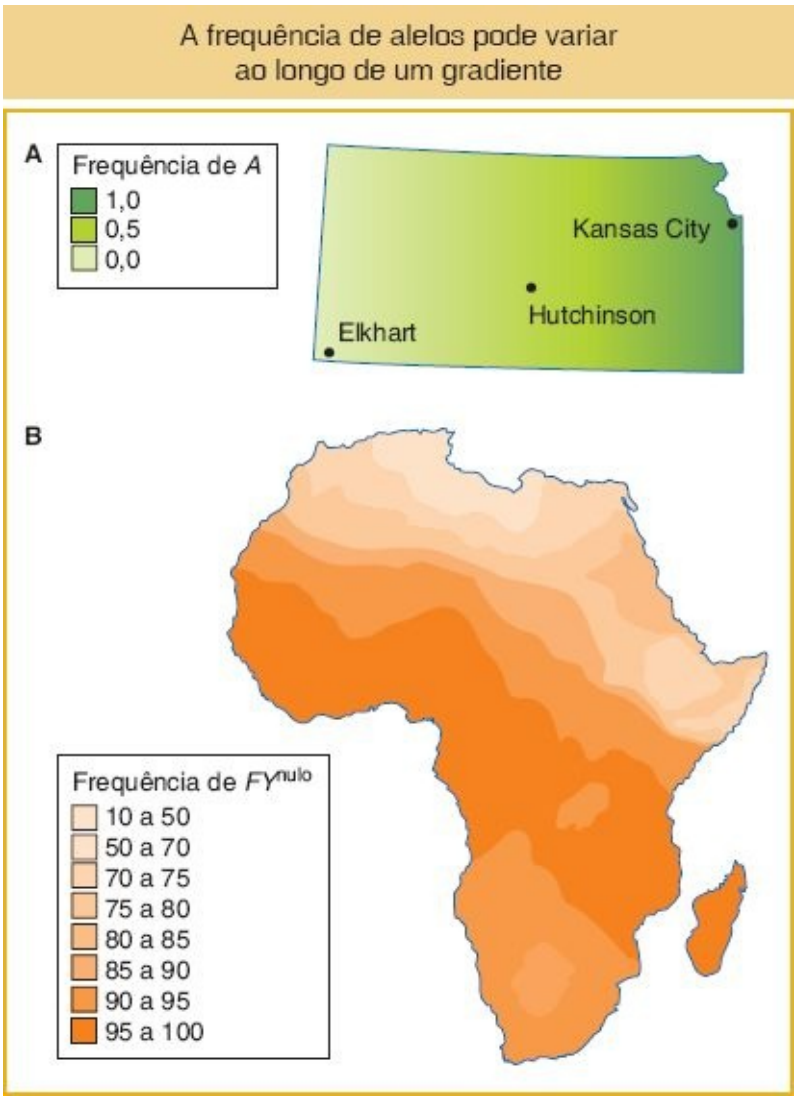


FIGURA 18.11 A. Variação da frequência alélica em todo o Kansas em relação a uma espécie hipotética de girassol selvagem. **B.** Variação da frequência em relação ao alelo FY^{nulo} do *locus* do grupo sanguíneo Duffy na África. (Dados de P. C. Sabeti *et al.*, *Science* 312, 2006, 1614-1620.)

> CONCEITO-CHAVE O cruzamento preferencial e o isolamento pela distância violam a lei de Hardy-Weinberg e podem fazer com que as frequências genotípicas desviem-se das expectativas de Hardy-Weinberg.

Endocruzamento

O terceiro tipo de desvio no acasalamento é o **endocruzamento**, ou cruzamento

entre parentes. Muito antes que qualquer pessoa soubesse a respeito dos alelos recessivos deletérios, algumas sociedades reconheciam que distúrbios como mudez, surdez e cegueira eram mais comuns entre os filhos de casamentos consanguíneos. Por isso, os casamentos de irmãos e irmãs e primos em primeiro grau são ilegais ou desencorajados. Apesar disso, muitos indivíduos famosos se casaram com primos, incluindo Charles Darwin, Albert Einstein, J. S. Bach, Edgar Allan Poe, Jesse James e a Rainha Victoria. Conforme veremos, a descendência dos casamentos consanguíneos tem risco mais alto de apresentar um distúrbio hereditário.

A progênie de endocruzamento apresenta maior probabilidade de ser homozigota em qualquer *locus* do que a progênie de cruzamentos não consanguíneos. Portanto, ela apresenta maior probabilidade de ser homozigota em relação a alelos recessivos deletérios. Por esse motivo, o endocruzamento pode levar a uma redução no vigor e no sucesso reprodutivo, denominada **depressão por endocruzamento**. Entretanto, o endocruzamento também apresenta vantagens. Em muitas espécies de plantas observa-se alta incidência de autopolinização e endogamia. Essas incluem a planta-modelo *Arabidopsis*, uma erva daninha de sucesso, e os cultivos de cereais produtivos de arroz e de trigo. Tendo em vista que a maior parte das espécies de plantas contém órgãos masculinos e femininos no mesmo indivíduo, a autopolinização pode ser conquistada mais facilmente do que o cruzamento externo. Outra vantagem da autopolinização é que, quando uma única semente é dispersa em um novo local, a planta que cresce a partir da semente apresenta um parceiro pronto — ela própria, o que possibilita que uma nova população seja estabelecida a partir de uma única semente. Finalmente, se uma planta individual apresenta uma combinação benéfica de alelos em diferentes *loci*, o endocruzamento então preserva aquela combinação. Nas espécies de plantas endogâmicas, benefícios como esses oferecem vantagens que superam o custo associado à depressão por endocruzamento.

Coeficiente de endocruzamento

O endocruzamento aumenta o risco de que um indivíduo seja homozigoto em relação a um alelo recessivo deletério e exiba uma doença genética. O aumento

desse risco depende de dois fatores: (1) frequência do alelo deletério na população e (2) grau de endocruzamento. Para medir o grau de endocruzamento, os geneticistas utilizam o **coeficiente de endocruzamento** (F), que é a probabilidade de que dois alelos em um indivíduo remontem à mesma cópia em um ancestral comum. Consideraremos primeiramente como calcular *F* com a utilização de heredogramas e em seguida examinaremos como *F* pode ser utilizado para determinar o aumento no risco de herdar uma doença recessiva.

Considere um heredograma simples em relação a um cruzamento entre meios-irmãos, indivíduos que apresentam um genitor em comum (Figura 18.12 A). Na figura, B e C são meios-irmãos que apresentam a mesma mãe, A, mas pais diferentes; B e C têm uma filha, I. Observe que existe uma alça fechada de I para B e A e de volta para I por C. A presença de uma alça fechada no heredograma nos informa que I é resultante de uma união consanguínea. As duas cópias do gene em A estão coloridas em azul e cor-de-rosa — o azul do pai de A e o cor-de-rosa de sua mãe. Conforme desenhado, I herdou a cópia cor-de-rosa tanto através de seu pai (B) quanto de sua mãe (C). Tendo em vista que as duas cópias do gene de I remontam à mesma cópia em sua avó, as suas duas cópias são **idênticas por descendência (IBD)**. Em termos mais gerais, se duas cópias de um gene em um indivíduo remontam à mesma cópia em um ancestral, então as cópias são IBD. Gostaríamos de ter um modo para calcular a probabilidade de que os dois alelos de I sejam IBD. Essa probabilidade é o coeficiente de endocruzamento para I, que é, na forma de símbolo, F_I.

Primeiramente, tendo em vista que estamos interessados apenas em traçar o trajeto dos alelos IBD, podemos simplificar o heredograma para que ele contenha apenas os indivíduos na alça fechada e ainda seguir a transmissão de quaisquer alelos IBD (Figura 18.12 B). Além disso, tendo em vista que o sexo do indivíduo não importa, utilizamos círculos para ambos os sexos. Os alelos transmitidos a cada cruzamento são rotulados *w*, *x*, *y* e *z*. Utilizamos "~" para simbolizar IBD. Gostaríamos de calcular a probabilidade de que *w* e *x* sejam IBD, mas faremos esse cálculo passo a passo. Primeiramente, qual é a probabilidade de que *x* e *y* sejam IBD ou, simbolicamente, qual é $P(x \sim y)$? Essa é a probabilidade de C transmitir a cópia herdada de A para I, que é 1/2, ou $P(x \sim y) = 1/2$. De modo

semelhante, a probabilidade de B transmitir a cópia herdada de A para I é 1/2, ou $P(w \sim z) = 1/2$.

Agora precisamos calcular a probabilidade de que *z* e *y* sejam IBD. Existem dois modos por meio dos quais *z* e *y* podem ser IBD. O primeiro modo é quando *z* e *y* são ambos a mesma cópia (ambas cor-de-rosa ou ambas azuis). Isso ocorre em 1/2 das ocasiões, tendo em vista que em 1/4 das ocasiões elas são ambas azuis e em 1/4, ambas cor-de-rosa. O segundo modo é quando *z* e *y* são cópias diferentes (uma cor-de-rosa e a outra azul), mas o indivíduo A foi endocruzado. Se o indivíduo A é endocruzado, existe então uma probabilidade de que as suas duas cópias do gene sejam IBD. A probabilidade de que as duas cópias de A sejam IBD é o coeficiente de endocruzamento de A, F_A. A probabilidade de que *z* e *y* sejam cópias diferentes (uma cor-de-rosa, a outra azul) é de 1/2. Assim, a probabilidade de que *z* e *y* sejam cópias diferentes que são IBD é 1/2, multiplicada pelo coeficiente de endocruzamento (F_A) para fornecer $\frac{1}{2}\ F_A$. Em conjunto, a probabilidade de que *z* e *y* sejam IBD é a probabilidade de que eles sejam a mesma cópia (1/2) mais a probabilidade de que eles sejam cópias diferentes que são IBD ($\frac{1}{2}\ F_A$). Simbolicamente, escrevemos:

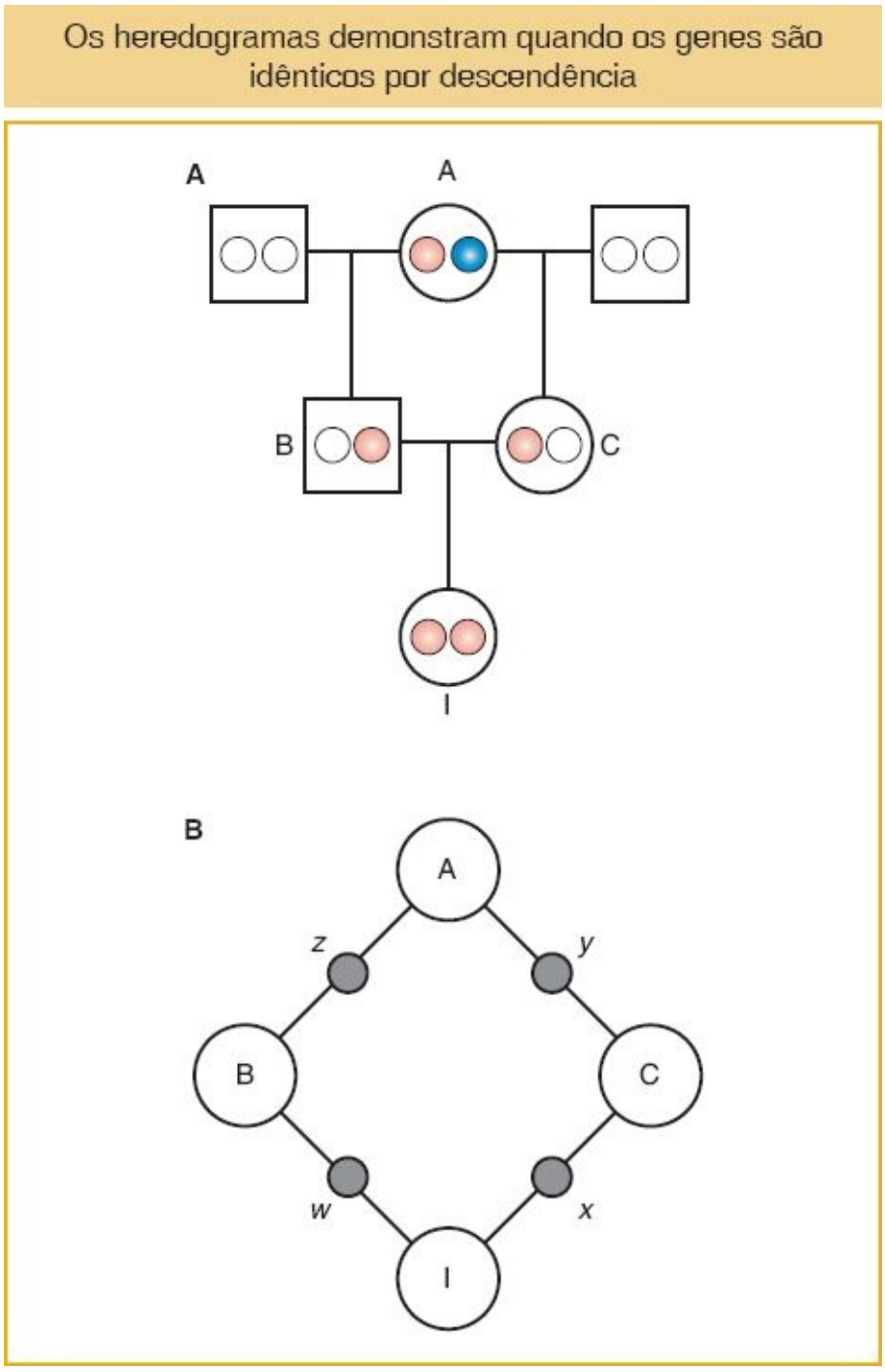


FIGURA 18.12 A. Heredograma em relação a um cruzamento entre meios-irmãos desenhado no formato-padrão. As *pequenas bolas coloridas* representam uma única cópia de um gene. No indivíduo A, as cópias *cor-de-rosa* e *azul* representam as cópias do gene que ela herdou de sua mãe e de seu pai, respectivamente. **B.** Heredograma em relação a um cruzamento entre meios-irmãos desenhado no formato simplificado utilizado para a análise de endocruzamento. Apenas as linhas que conectam o genitor à descendência estão desenhadas e apenas os indivíduos na "alça de endocruzamento fechada" estão incluídos. *w*, *x*, *y* e *z* são símbolos para o alelo transmitido do genitor para a descendência.

$$P(z \sim y) = + F_A$$

P(x ~ y), *P(w ~ z)* e *P(z ~ y)* são probabilidades independentes; portanto, podemos utilizar a regra do produto e reunir tudo para obter:

$F_I = P(x \sim y) \times P(w \sim z) \times P(z \sim y)$
$= \frac{1}{2} \times \frac{1}{2} \times (\frac{1}{2} + \frac{1}{2} F_A)$
$= (\frac{1}{2})^3 (1 + F_A)$

Na análise de heredogramas consanguíneos, podemos substituir o valor de F_A na equação anterior, se ele for conhecido. De outro modo, podemos presumir que F_A é zero, se não houver informação para sugerir que o indivíduo A seja endocruzado. No exemplo atual, se presumirmos que $F_A = 0$, então:

$$F_I = (\frac{1}{2})^3 = \frac{1}{8}$$

Esse cálculo nos informa que a descendência de cruzamentos entre meios-irmãos será homozigota em relação aos alelos que são IBD em relação a no mínimo 1/8 de seus genes. Ela pode ser superior a 1/8 se F_A for superior a zero. Heredogramas de endocruzamentos adicionais e uma fórmula geral para o cálculo de *F* podem ser encontrados no Quadro 18.2.

Quando existe um endocruzamento em uma população, a presunção de cruzamento aleatório de Hardy-Weinberg será violada. Entretanto, a lei de Hardy-Weinberg pode ser modificada para corrigir as proporções genotípicas previstas em relação a diferentes graus de endocruzamento por meio da utilização de *F*, o coeficiente de endocruzamento médio para a população. As frequências de Hardy-Weinberg modificadas são:

$f_{A/A} = p^2 + pqF$
$f_{A/a} = 2pq — 2pqF$
$f_{a/a} = q^2 + pqF$

Essas proporções de Hardy-Weinberg modificadas fazem sentido intuitivamente, demonstrando como o endocruzamento reduz a frequência de heterozigotos em

$2pqF$ e adiciona metade dessa quantidade a cada uma das classes homozigotas. Com essas equações de Hardy-Weinberg modificadas, você também observará que quando não há endocruzamento ($F = 0$), você obtém novamente as frequências genotípicas de Hardy-Weinberg padrão e, quando existe um endocruzamento completo ($F = 1$), você obtém $f_{A/A} = p$ e $f_{a/a} = q$.

Em quanto o endocruzamento aumenta o risco de que aquela descendência exiba uma condição de doença recessiva? A Tabela 18.3 demonstra os coeficientes de endocruzamento em relação à descendência e alguns diferentes cruzamentos consanguíneos e o número previsto de homozigotos recessivos em relação a diferentes frequências (q) do alelo recessivo. Quando $q = 0{,}01$, existe um aumento de 7 vezes (7,19/1,0) na descendência homozigota recessiva para cruzamentos entre primos de primeiro grau, em comparação aos cruzamentos entre indivíduos não relacionados. O aumento no risco aumenta para 13 vezes (3,36/0,25) quando $q = 0{,}005$ e para 63 vezes (0,63/0,01) quando $q = 0{,}001$. Em outras palavras, o risco aumenta dramaticamente em relação aos alelos raros. Cruzamentos entre irmão e irmã e entre genitores e filhos são os mais arriscados: quando $q = 0{,}001$, eles demonstram um risco 250 vezes (2,51/0,01) maior em comparação aos cruzamentos entre indivíduos não relacionados.

Quadro 18.2 Cálculo dos coeficientes de endocruzamento a partir de heredogramas.

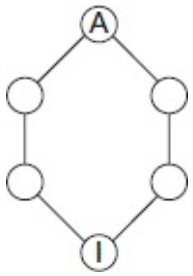
No texto, vimos que o coeficiente de endocruzamento (F_I) em relação à descendência de um cruzamento entre meios-irmãos é:

$$F_I = \left(\tfrac{1}{2}\right)^3 (1 + F_A)$$

em que F_A é o coeficiente de endocruzamento do ancestral. Essa expressão inclui o termo 1/2 elevado à terceira potência, $(½)^3$. Na Figura 18.12, você verá que existem três indivíduos na alça de endocruzamento, sem contar I. A fórmula geral para computar os coeficientes de endocruzamento a partir de heredogramas é:

$$F_I = (\tfrac{1}{2})^n (1 + F_A)$$

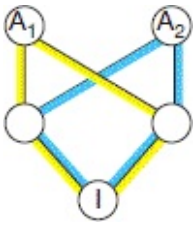
em que *n* é o número de indivíduos na alça de endocruzamento, sem contar I. Vejamos outro heredograma, no qual os avós de I são meios-irmãos:



Existem cinco indivíduos na alça de endocruzamento além de I; assim, se presumirmos que o ancestral não foi endocruzado ($F_A = 0$), então:

$$F_I = (\tfrac{1}{2})^5(1 + F_A) = 0{,}03125$$

Em alguns heredogramas, existe mais de uma alça de endocruzamento. Aqui está um heredograma no qual I é a descendência de um cruzamento entre irmãos legítimos:



Em relação aos heredogramas com diversas alças de endocruzamento, soma-se a contribuição em todas as alças em que F_A é o coeficiente de endocruzamento do ancestral (*A*) da alça em questão:

$$F_I = \sum_{\text{alças}} (\tfrac{1}{2})^n (1 + F_A)$$

Portanto, em relação ao heredograma no qual I é a descendência de um cruzamento entre irmãos legítimos, obtemos:

$$F_I = (\tfrac{1}{2})^3(1 + F_{A1}) + (\tfrac{1}{2})^3 (1 + F_{A2}) = \tfrac{1}{4}$$

presumindo que os coeficientes de endocruzamento em relação a ambos os ancestrais sejam 0.

Tabela 18.3 Número de homozigotos recessivos por 10.000 indivíduos em relação a diferentes frequências alélicas (*q*).

Cruzamento	***F***	***q* = 0,01**	***q* = 0,005**	***q* = 0,001**
Genitores não relacionados	0,0	1,00	0,25	0,01
Genitor e filho ou irmão e irmã	1/4	25,75	12,69	2,51
Meios-irmãos	1/8	13,38	6,47	1,26
Primos de primeiro grau	1/16	7,19	3,36	0,63
Primos de segundo grau	1/64	2,55	1,03	0,17

O impacto do endocruzamento sobre a frequência de distúrbios genéticos em populações humanas pode ser observado na Figura 18.13. Os filhos de casamentos de primos em primeiro grau demonstram uma frequência duas vezes mais alta de distúrbios em comparação aos filhos de genitores não relacionados.

Registros históricos sugerem que os riscos do endocruzamento eram compreendidos muito antes da existência da genética.

Tamanho da população e endocruzamento

O tamanho da população é um fator de contribuição importante para o nível de endocruzamento nas populações. Em populações pequenas, os indivíduos apresentam maior probabilidade de cruzar com um parente do que em populações maiores. O fenômeno é observado em pequenas populações humanas, tais como a das Ilhas Tristão da Cunha, no Atlântico Sul, que tem menos de 300 pessoas. Vejamos o efeito do tamanho da população sobre o nível geral de endocruzamento em uma população, conforme medido por *F*.

Considere uma população com F_t sendo o nível de endocruzamento na geração *t*. Para formar um indivíduo na próxima geração $t + 1$, selecionamos o primeiro alelo do *pool* gênico. Suponha que o tamanho da população é *N*. Após a seleção do primeiro alelo, a probabilidade de que o segundo alelo que coletamos seja exatamente a mesma cópia é $1/2N$ e o coeficiente de endocruzamento em relação a esse indivíduo é 1,0. A probabilidade de que o segundo alelo que coletamos seja uma cópia diferente do primeiro alelo é $1 – 1/2N$ e o nível de endocruzamento em relação ao indivíduo resultante seria F_t, o coeficiente de endocruzamento médio em relação à população inicial na geração *t*. O nível de endocruzamento na próxima geração é a soma desses dois desfechos possíveis, ou:

$$F_{t+1} = \left(\frac{1}{2N}\right)1 + \left(1 - \frac{1}{2N}\right)F_t$$

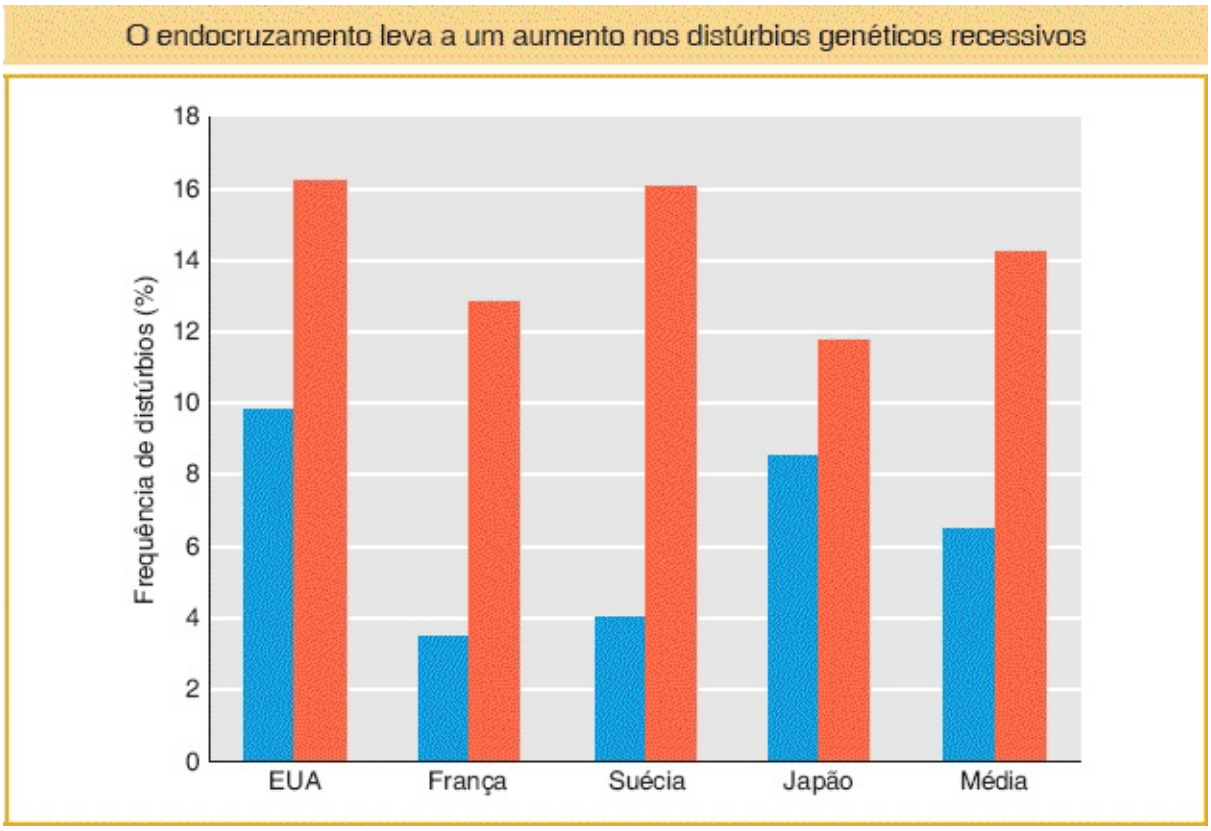


FIGURA 18.13 A frequência de distúrbios genéticos entre filhos de genitores não relacionados (*colunas azuis*) em comparação àquela de filhos de genitores que são primos de primeiro grau (*colunas vermelhas*). (Dados de C. Stern, *Principles of Human Genetics*, W. H. Freeman, 1973.)

Essa equação nos informa que a F aumentará ao longo do tempo como uma função do tamanho da população. Quando N é grande, F aumenta lentamente ao longo do tempo. Quando N é pequeno, F aumenta rapidamente ao longo do tempo. Por exemplo, suponha que F_t na população inicial seja 0,1 e N = 10.000. Em seguida, F_{t+1} seria 0,10005, um valor apenas ligeiramente mais alto. Entretanto, se N = 10, então F_{t+1} seria 0,145, um valor muito mais alto. Também podemos utilizar essa equação recursivamente para calcular F_{t+2} por meio da utilização de F_{t+1} em vez de F_t do lado direito. O resultado com N = 10 e F_t = 0,1 será F_{t+2} = 0,188. Os efeitos do tamanho da população sobre o endocruzamento em populações são adicionalmente explorados no Quadro 18.3.

Uma consequência do aumento de endocruzamento é que os indivíduos em populações pequenas apresentam maior probabilidade de ser homozigotos em relação a alelos deletérios, assim como a descendência de casamentos entre

primos em primeiro grau apresenta maior probabilidade de ser homozigota em relação a tais alelos. Esse efeito é observado em grupos étnicos que vivem em pequenas comunidades isoladas em termos reprodutivos. Por exemplo, um tipo de nanismo no qual os indivíduos afetados apresentam seis dedos ocorre com uma frequência superior a 1 em 200 em uma população de aproximadamente 13.000 Amish no Condado de Lancaster, na Pensilvânia, embora sua frequência na população norte-americana em geral seja de apenas 1 em 60.000.

Quadro 18.3 Endocruzamento em populações finitas.

No texto principal, derivamos a fórmula em relação ao aumento no endocruzamento entre gerações em populações finitas como:

$$F_{t+1} = \left(\frac{1}{2N}\right)1 + \left(1 - \frac{1}{2N}\right)F_t$$

que pode ser reescrita como:

$$(1 - F_{t+1}) = \left(1 - \frac{1}{2N}\right)(1 - F_t)$$

Também apresentamos a fórmula em relação à frequência de heterozigotos (H) com endocruzamento como:

$$H = f_{A/a} = 2pq - 2pqF$$

que pode ser reescrita como:

$$(1 - F) = H/2pq$$

Combinando essas duas equações, obtemos:

$$H_{t+1}/2pq = \left(1 - \frac{1}{2N}\right)H_t/2pq$$

e em seguida:

$$H_{t+1} = \left(1 - \frac{1}{2N}\right) H_t$$

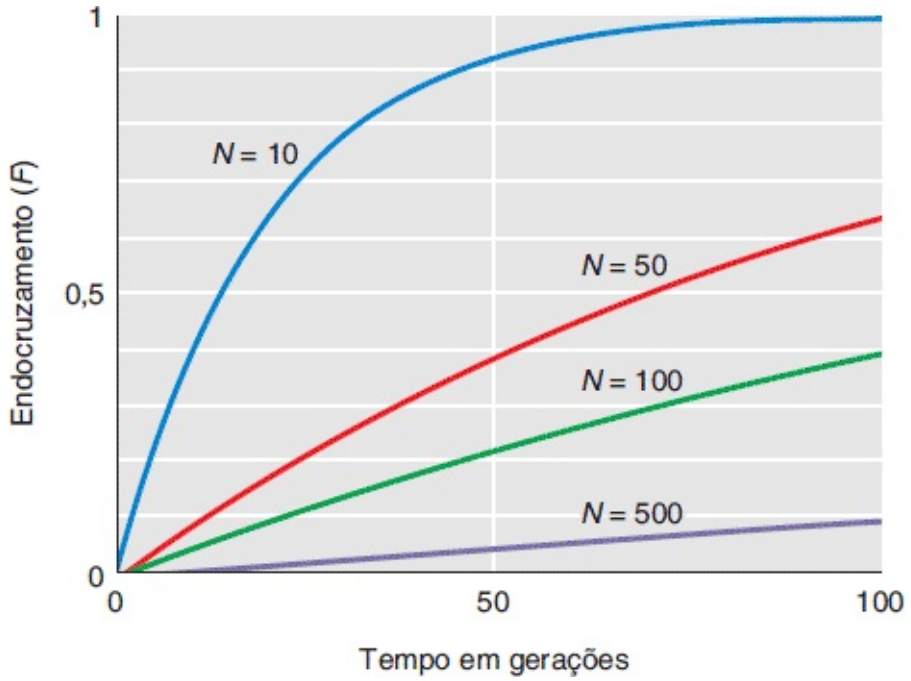
Portanto, em relação a cada geração, o nível de heterozigosidade é reduzido pela fração (1 — 1/2*N*). A redução em *H* ao longo de *t* gerações é:

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0$$

e a alteração em *F* ao longo das gerações *t* é fornecida por:

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t (1 - F_0)$$

Conforme demonstrado na figura a seguir, o endocruzamento aumentará com o tempo em uma população finita, até mesmo quando não houver endocruzamento na população inicial.



Aumento no endocruzamento (*F*) ao longo do tempo em relação a diversos tamanhos de populações diferentes.

> CONCEITO-CHAVE O endocruzamento aumenta a frequência de homozigotos em uma população, e pode resultar em uma frequência mais alta de distúrbios genéticos recessivos. O coeficiente de endocruzamento (*F*) é a probabilidade de que dois alelos em um indivíduo remontem à mesma cópia em um ancestral comum.

18.4 Variação genética e sua medida

Para estudar a quantidade e a distribuição da variação genética em populações, precisamos de alguns modos para quantificar a variação. Para descrever como podemos quantificar a variação, utilizaremos dados em relação ao gene da glicose-6-fosfato desidrogenase (*G6PD*) de seres humanos. O *G6PD* é um gene ligado ao X que codifica uma enzima que catalisa uma etapa na glicólise. O alelo do tipo selvagem (*B*) de *G6PD* apresenta atividade enzimática total. Um segundo alelo, denominado A^{-}, leva à atividade enzimática fortemente reduzida, e os indivíduos que carregam esse alelo desenvolvem anemia hemolítica. Entretanto, esse alelo também confere uma redução de 50% no risco de malária grave nos portadores. Em regiões da África nas quais a malária é endêmica, o alelo A^{-} alcança frequências próximas de 20%, embora esse alelo esteja ausente ou seja raro em outros locais. Outro alelo (A^{+}) leva a uma atividade enzimática apenas modestamente reduzida. Contrariamente aos indivíduos que carregam o alelo A^{-}, os indivíduos que carregam apenas os alelos A^{+} ou *B* não desenvolvem anemia hemolítica.

A Figura 18.14 demonstra os SNP em 18 sítios polimórficos que foram identificados por meio do sequenciamento de um segmento de 5.102 pb do *G6PD* a partir de uma amostra mundial de 47 homens. Os 5.084 sítios remanescentes eram **fixado**, ou invariantes: existe apenas um único alelo (nucleotídio) na amostra inteira em relação a cada um desses sítios. Ao fazer amostragem apenas com homens, observamos apenas um alelo e um haplótipo para cada indivíduo, tendo em vista que o gene é ligado ao X. O alelo A^{+} difere de *B* pela substituição de um único aminoácido (ácido aspártico em lugar de asparagina) no *SNP3* na

Figura 18.14. O alelo A^- difere do alelo *B* em dois aminoácidos: ele contém ambas a substituição do "ácido aspártico em lugar da asparagina" observada no alelo A^+ e uma diferença em um segundo aminoácido (metionina em lugar de valina) no *SNP2*.

Como podemos quantificar a variação no *locus G6PD*? Uma medida simples é o número de **sítios segregantes (*S*)** ou polimórficos. Em relação aos dados de *G6PD*, *S* é 18 na amostra total, 14 na amostra africana e 7 na amostra não africana. Os africanos contêm o dobro do número de sítios segregantes, apesar do fato de nossa amostra apresentar menos africanos. Outra medida simples é o **número de haplótipos (*NH*)**. O valor da *NH* é 12 na amostra total, 9 na amostra africana e 6 na amostra não africana. Novamente, a amostra africana apresenta maior variação. Uma falha de medidas tais como *S* e *NH* é que os valores que observamos dependem fortemente do tamanho da amostra. Se temos amostras com mais indivíduos, então os valores de *S* e *NH* tendem a aumentar. Por exemplo, nossa amostra apresenta 16 africanos em comparação a 31 não africanos. Embora *S* tenha o dobro do tamanho em africanos do que em não africanos, a diferença provavelmente seria ainda maior se apresentássemos um número igual (31) de africanos e não africanos.

Em lugar de *S* e *NH*, podemos calcular as frequências alélicas, que não são influenciadas por diferenças no tamanho da amostra. Em relação aos dados de *G6PD*, *B*, A^- e A^+ apresentam frequências mundiais de 0,83, 0,13 e 0,04, respectivamente. Entretanto, você observará que A^- apresenta uma frequência de 0,0 fora da África e de 0,38 em nossa amostra africana, que é uma diferença substancial. Podemos utilizar os dados da frequência alélica para calcular uma estatística denominada **diversidade gênica (*GD*)**, que é a probabilidade de que dois alelos coletados aleatoriamente do *pool* gênico sejam diferentes. A probabilidade de coleta de dois alelos diferentes é igual a 1 menos a probabilidade de coleta de duas cópias do mesmo alelo somada a todos os alelos do *locus*. Portanto:

$$\begin{aligned} GD &= 1 - \sum p_i^2 \\ &= 1 - (p_1^2 + p_2^2 + p_3^2 + \ldots p_n^2) \end{aligned}$$

em que p_i é a frequência do $i^{ésimo}$ alelo e Σ é o sinal de somatório, indicando que adicionamos os quadrados de todos os *n* valores observados de *p* para *i* = 1, 2, até o $n^{ésimo}$ alelo. O valor de *GD* pode variar de 0 a 1. Ele se aproximará de 1 quando houver um grande número de alelos de frequências aproximadamente iguais. Será 0 quando houver um único alelo, e estará próximo de 0 sempre que houver um único alelo muito comum com uma frequência de 0,99 ou superior. A Tabela 18.4 demonstra que a diversidade gênica é razoavelmente alta em africanos (0,47). Tendo em vista que os não africanos apresentam apenas o alelo *B*, a diversidade gênica é 0,0.

O valor de *GD* é igual à proporção esperada de heterozigotos sob o equilíbrio de Hardy-Weinberg, **heterozigosidade (*H*)**. Entretanto, *H* como um conceito é aplicável apenas aos diploides e não se aplicaria aos *loci* ligados ao X em indivíduos do sexo masculino. Portanto, conceitualmente a diversidade gênica (*GD*) é mais apropriada até mesmo se for matematicamente a mesma quantidade de *H* para populações de diploides sob o equilíbrio de Hardy-Weinberg.

A diversidade gênica pode ser calculada em relação a um sítio de nucleotídio único. Ela pode ser calculada para todos os sítios de nucleotídios em um gene, caso em que é denominada **diversidade nucleotídica**. Tendo em vista que a maioria dos nucleotídios em quaisquer duas cópias de um gene de uma espécie é tipicamente a mesma, os valores para a diversidade nucleotídica em relação aos genes são tipicamente muito pequenos. Em relação à *G6PD*, existem apenas 18 sítios de nucleotídios polimórficos, mas 5.084 sítios invariantes. A diversidade nucleotídica média para toda a sequência do gene *G6PD* é de 0,0008 em africanos, 0,0002 em não africanos, e 0,0006 na amostra total. Esses valores nos informam que os africanos apresentam quatro vezes mais diversidade nucleotídica em *G6PD* do que os não africanos.

Tabela 18.4 Dados de diversidade em relação à glicose-6-fosfato desidrogenase (*G6PD*) em seres humanos.

	Amostra total	Africanos	Não africanos

Tamanho da amostra	47	16	31
Número de sítios segregantes	18	14	7
Número de haplótipos	12	9	6
Diversidade gênica (*GD*) no *SNP2*	0,22	0,47	0,00
Diversidade de nucleotídios	0,0006	0,0008	0,0002

Variação de nucleotídios no gene *G6PD* em seres humanos

Indivíduo	Origem	Alelo	SNP																		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Haplótipo
			A	G	A	C	C	G	C	C	C	C	C	G	G	C	T	C	A	C	
1	África Austral	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
2	África Central	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
3	África Central	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
4	Afro-americano	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
5	Afro-americano	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
6	África Central	*A-*	G	A	G	•	G	•	•	T	•	T	•	•	•	•	C	•	G	•	1
7	África Central	*A+*	G	•	G	•	•	•	•	•	•	T	•	•	•	•	C	•	G	•	2
8	África Central	*A+*	G	•	G	•	•	•	•	•	•	T	•	•	•	•	C	•	G	•	2
9	África Central	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C	•	G	•	3
10	África Austral	*B*	•	•	•	•	•	•	•	•	•	•	•	A	•	•	C	T	G	•	4
11	África Austral	*B*	•	•	•	•	•	•	•	•	•	•	•	A	•	•	C	T	G	•	4
12	África Austral	*B*	•	•	•	T	•	•	•	•	•	•	•	•	•	•	C	T	G	•	5
13	África Austral	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C	T	G	•	6
14	África Austral	*B*	•	•	•	•	•	•	•	•	•	•	T	•	A	•	C	•	G	•	7
15	África Central	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	T	C	•	G	•	8
16	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	T	C	•	G	•	8
17	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	T	C	•	G	•	8
18	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	T	C	•	G	•	8
19	Sudoeste Asiático	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	T	C	•	G	•	8
20	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C	•	G	•	3
21	Nativo americano	*B*	•	•	•	•	•	A	T	•	•	•	•	•	•	•	C	•	G	•	9
22	África Austral	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
23	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
24	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
25	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
26	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
27	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
28	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
29	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
30	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
31	Nativo-americano	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
32	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
33	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
34	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
35	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
36	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
37	Europa	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
38	Sudoeste Asiático	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
39	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
40	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
41	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
42	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
43	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
44	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
45	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	10
46	Ilhas do Pacífico	*B*	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•	•	•	•	11
47	Ásia Oriental	*B*	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	12

FIGURA 18.14 Variação de nucleotídios de 5.102 pb do gene *G6PD* em relação a uma amostra mundial de 47 homens. Apenas os 18 sítios variáveis estão demonstrados. A classe alélica funcional (A^{-}, A^{+}, ou *B*) está

demonstrada em relação a cada sequência. O *SNP2* é um SNP não sinônimo que causa uma substituição de valina para metionina, que é causa das diferenças na atividade enzimática associada ao alelo A^{-}. O *SNP3* é um SNP não sinônimo que causa substituição do aminoácido ácido aspártico por asparagina. (Dados de M. A. Saunders *et al.*, *Genetics* 162, 2002, 1849-1861.)

A Figura 18.15 demonstra o nível de diversidade nucleotídica em diversos organismos. Os eucariotos unicelulares são os mais diversos, seguidos pelas plantas e em seguida pelos invertebrados. Os vertebrados são o grupo menos diverso; entretanto, a maior parte dos vertebrados ainda apresenta muita diversidade nucleotídica. Em relação aos seres humanos, a diversidade nucleotídica é de aproximadamente 0,001, o que significa que dois cromossomos humanos escolhidos aleatoriamente diferirão em aproximadamente 1 pb por mil. Com 3 bilhões de pb em nosso genoma, isso soma um total de aproximadamente 3 milhões de diferenças entre o conjunto de cromossomos herdado da mãe de uma pessoa e o conjunto herdado do pai de uma pessoa para os indivíduos não consanguíneos.

> CONCEITO-CHAVE As populações biológicas com frequência são ricas em variação genética. Essa diversidade pode ser quantificada por meio de diferentes estatísticas para comparar os níveis de variação entre as populações e as espécies.

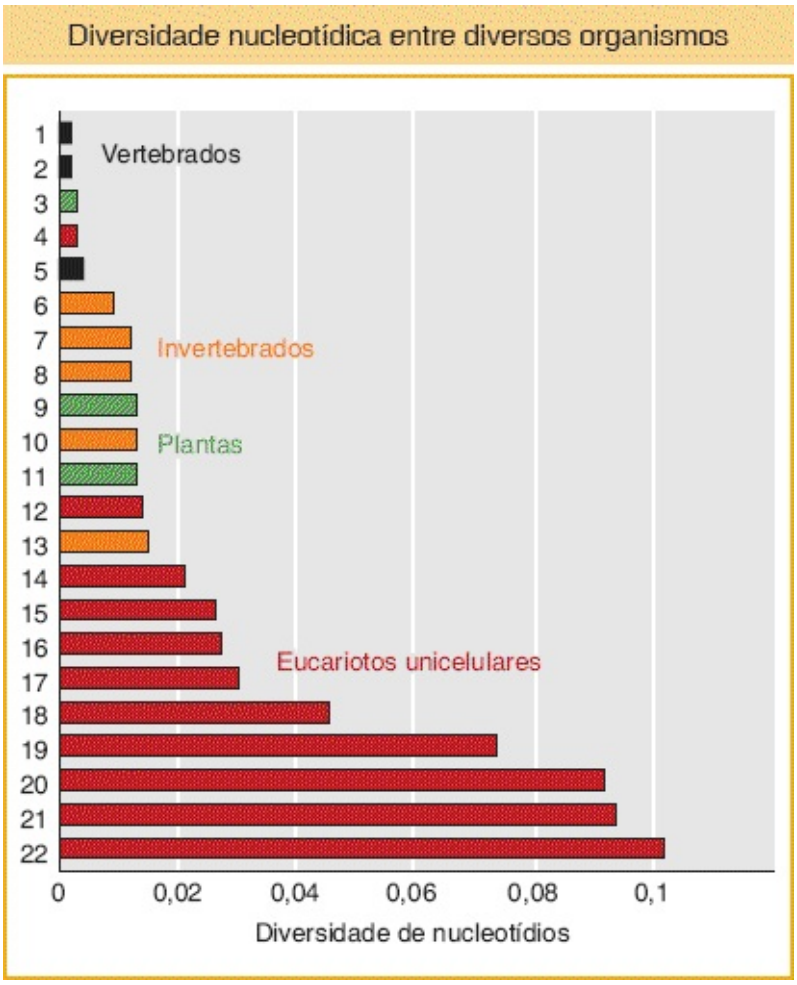


FIGURA 18.15 Níveis de diversidade nucleotídica em sítios sinônimos e silenciosos em alguns organismos diferentes. (1) *Mus musculus*, (2) *Homo sapiens*, (3) *Oryza sativa*, (4) *Plasmodium falciparum*, (5) *Fugu rubripes*, (6) *Strongylocentrotus purpuratus*, (7) *Anopheles gambiae*, (8) *Ciona intestinalis*, (9) *Arabidopsis thaliana*, (10) *Caenorhabditis elegans*, (11) *Zea mays*, (12) *Encephalitozoon cuniculi*, (13) *Drosophila melanogaster*, (14) *Leishmania major*, (15) espécies de *Trypanosoma*, (16) *Toxoplasma gondii*, (17) *Giardia lamblia*, (18) *Neurospora crassa*, (19) *Dictyostelium discoideum*, (20) *Saccharomyces cerevisiae*, (21) *Cryptosporidium parvum*, (22) *Cryptococcus neoformans*. (Dados de M. Lynch e J. S. Conery, *Science* 302, 2003, 1401-1404.)

18.5 Modulação da variação genética

Quais são as forças que modulam a quantidade de variação genética em uma população? Como os novos alelos entram no *pool* gênico? Quais forças removem alelos do *pool* gênico? Como as variantes genéticas podem ser recombinadas para criar novas combinações de alelos? As respostas a essas questões são o

cerne da compreensão do processo da evolução. Nesta seção, examinaremos os papéis da mutação, da migração, da recombinação, da deriva genética (acaso) e da seleção na formação da composição genética das populações.

Novos alelos na população | Mutação e migração

A mutação é a fonte definitiva de toda a variação genética. No Capítulo 16, discutimos os mecanismos moleculares que são a base das mutações em pequena escala, tais como mutações de ponto, indels e alterações no número de unidades de repetição em microssatélites. Os geneticistas de populações estão particularmente interessados na **taxa de mutação**, que é a probabilidade de que uma cópia de um alelo seja alterada para alguma outra forma alélica em uma geração. A taxa de mutação é tipicamente simbolizada pela letra grega **μ**. Conforme veremos a seguir, se soubermos a taxa de mutação e o número de diferenças nucleotídicas entre duas sequências, então poderemos estimar há quanto tempo as duas sequências divergiram.

Como os geneticistas conseguem estimar a taxa de mutação? Os geneticistas podem estimar as taxas de mutação iniciando com um único indivíduo homozigoto e seguindo o heredograma de seus descendentes durante diversas gerações. Então eles podem comparar a sequência de DNA do indivíduo fundador com as sequências de DNA dos descendentes diversas gerações depois e registrar quaisquer novas mutações que tenham ocorrido. O número de mutações observadas por genoma por geração fornece uma estimativa da taxa. Tendo em vista que estamos procurando eventos um tanto quanto raros, é necessário sequenciar bilhões de nucleotídios para encontrar apenas algumas mutações de SNP. Em 2009, a taxa de mutação de SNP em relação a uma parte do cromossomo Y humano foi estimada por meio dessa abordagem como sendo de $3{,}0 \times 10^{-8}$ mutações/nucleotídio/geração, ou aproximadamente uma mutação a cada 30 milhões de pb. Se extrapolarmos para o genoma humano inteiro (3 bilhões de pb), cada um de nós herdou então 100 novas mutações de cada um de nossos genitores. Felizmente, a maioria das mutações não é prejudicial, tendo em vista que elas ocorrem em regiões do genoma que não são críticas.

A Tabela 18.5 lista as taxas de mutação em relação a SNP e microssatélites em

diversos organismos-modelo. A taxa de mutação de SNP é diversas ordens de magnitude mais baixa do que a taxa de microssatélites. A sua mais alta taxa de mutação e a maior variação tornam os microssatélites particularmente úteis na genética de populações e na ciência forense do DNA. A taxa de mutação de SNP por geração aparenta ser mais baixa em relação a organismos unicelulares do que em relação a organismos multicelulares. Essa diferença pode ser explicada no mínimo parcialmente pelo número de divisões celulares por geração. Existem aproximadamente 200 divisões celulares desde o zigoto até o gameta em seres humanos, mas apenas uma em *E. coli*. Se a taxa humana for dividida por 200, então a taxa por divisão celular em seres humanos está notavelmente próxima da taxa em *E. coli*.

Tabela 18.5 Taxas de mutação aproximadas por geração por genoma haploide.

Organismo	Mutações de SNP (por pb)	Microssatélite
Arabidopsis	7×10^{-9}	9×10^{-4}
Milho	3×10^{-8}	8×10^{-4}
E. coli	5×10^{-10}	—
Levedura	5×10^{-10}	4×10^{-5}
C. elegans	3×10^{-9}	4×10^{-3}
Drosophila	4×10^{-9}	9×10^{-6}
Camundongo	4×10^{-9}	3×10^{-4}

Ser humano	3×10^{-8}	6×10^{-4}

Nota: a taxa de microssatélites é em relação a repetições de dinucleotídios ou trinucleotídios.
Fonte: dados de diversos estudos publicados.

Além da mutação, o único outro meio para a entrada de nova variação em uma população é por meio de **migração** ou **fluxo gênico**, a movimentação de indivíduos (ou gametas) entre as populações. A maior parte das espécies é dividida em um conjunto de pequenas populações ou subpopulações locais. As barreiras físicas, tais como oceanos, rios ou montanhas, podem reduzir o fluxo gênico entre as subpopulações, mas com frequência ocorre algum grau de fluxo gênico, apesar de tais barreiras. Dentro das subpopulações, um indivíduo pode ter uma chance de acasalar com qualquer outro membro do sexo oposto; entretanto, os indivíduos de diferentes subpopulações não podem acasalar, exceto se houver migração.

As subpopulações isoladas tendem a divergir na medida em que cada uma acumula as suas próprias mutações únicas. O fluxo gênico limita a divergência genética entre as subpopulações. Uma das consequências genéticas da migração é a **mistura genética**, a mistura de genes que resulta quando os indivíduos apresentam antepassados de mais de uma subpopulação. Esse fenômeno é comum em populações humanas. Ele é prontamente observado na África do Sul, para onde migrantes de todo o mundo foram levados. Conforme demonstrado na Figura 18.16, os genomas dos sul-africanos de ancestralidade mista são complexos e incluem partes da população indígena da África Austral, mais contribuições de migrantes da África Ocidental, da Europa, da Índia, da Ásia Oriental e de outras regiões.

> CONCEITO-CHAVE A mutação é a fonte definitiva de toda a variação genética. A migração pode adicionar variação genética a uma população por meio do fluxo gênico de outra população da mesma espécie.

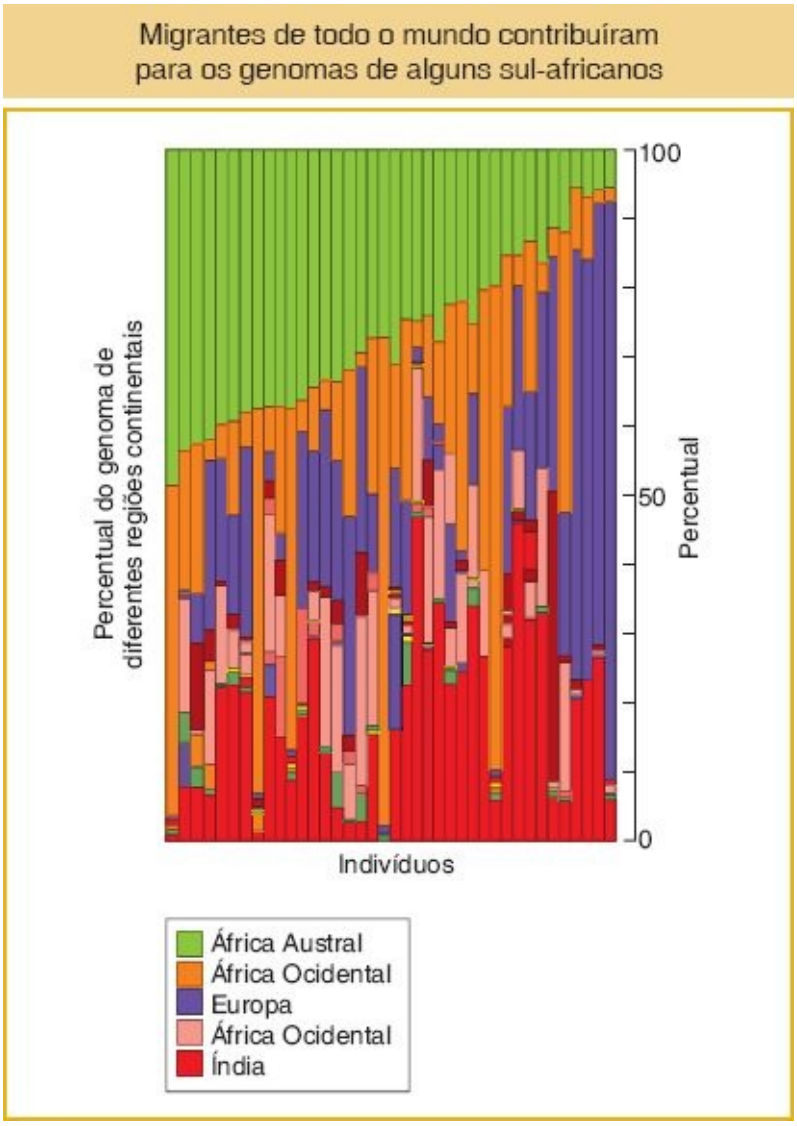


FIGURA 18.16 Representação gráfica da mistura genética em 39 pessoas de ancestralidade mista da África do Sul. Cada coluna representa o genoma de uma pessoa, e as cores representam as partes de seu genoma com contribuição por parte de seus ancestrais, que vieram de muitas regiões do mundo. A figura tem por base a análise genética de populações de mais de 800 microssatélites e 500 indels que foram classificados em quase 4.000 pessoas de todo o mundo, incluindo as 39 de ancestralidade mista da África do Sul. (Dados de S. A. Tishkoff *et al.*, *Science* 324, 2009, 1035-1044.)

Recombinação e desequilíbrio de ligação

A recombinação é uma força crítica que molda os padrões da variação genética nas populações. Nesse caso, não existe ganho ou perda de alelos; em vez disso, a

recombinação cria novos haplótipos. Vejamos como isso funciona. Considere os *loci A* e *B* ligados. Pode haver uma população na qual apenas dois haplótipos são observados na geração t_0: *AB* e *ab*. Suponha que um indivíduo nessa população é heterozigoto em relação a esses dois haplótipos:

$$\frac{A \qquad B}{a \qquad b}$$

Se ocorrer um *crossing over* nesse indivíduo, então gametas com dois novos haplótipos, *Ab* e *aB*, podem ser formados e entrar na população na geração t_1.

$$\underline{A \qquad B} \qquad \underline{a \qquad b}$$

Portanto, a recombinação pode criar a variação que adota a forma de novos haplótipos. Os novos haplótipos podem apresentar propriedades únicas que alterem a função proteica. Por exemplo, suponha que um aminoácido variante em uma proteína em um haplótipo aumente a atividade enzimática da proteína em duas vezes e que um segundo aminoácido variante em outro haplótipo também aumente a atividade em duas vezes. Um evento de recombinação que combine essas duas variantes produziria uma proteína com uma atividade quatro vezes mais alta.

Agora consideraremos as frequências observadas e esperadas dos quatro possíveis haplótipos em relação a dois *loci*, cada um com dois alelos. Os *loci* ligados, *A* e *B*, apresentam os alelos *A* e *a* e *B* e *b*, com as frequências p_A, p_a, p_B, e p_b, respectivamente. Os quatro haplótipos possíveis são *AB*, *Ab*, *aB*, e *ab*, com as frequências observadas P_{AB}, P_{Ab}, P_{aB} *e* P_{ab}. A qual frequência esperamos observar cada um desses quatro haplótipos? Se houver uma relação aleatória entre os alelos nos dois *loci*, então a frequência de qualquer haplótipo será o produto das frequências dos dois alelos que compõem aquele haplótipo:

$P_{AB} = p_A \times p_B$
$P_{Ab} = p_A \times p_b$
$P_{aB} = p_a \times p_B$

$P_{ab} = p_a \times p_b$

Por exemplo, suponha que a frequência de cada um dos alelos é de 0,5; ou seja, $p_A = p_a = p_B = p_b = 0{,}5$. Quando obtemos uma amostra do *pool* gênico, a probabilidade de coleta de um cromossomo com um alelo *A* é de 0,5. Se a relação entre os alelos no *locus A* e os alelos no *locus B* for aleatória, então a probabilidade de que o cromossomo selecionado apresente o alelo *B* também é de 0,5. Portanto, a probabilidade de coletarmos um cromossomo com o haplótipo *AB* é:

$$P_{AB} = p_A \times p_B = 0{,}5 \times 0{,}5 = 0{,}25$$

Se a associação entre os alelos em dois *loci* for aleatória, conforme descrito há pouco, diz-se então que os dois *loci* estão em **equilíbrio de ligação**. Nesse caso, as frequências observadas e esperadas serão as mesmas. A Figura 18.17 A diagrama um caso de dois *loci* em equilíbrio de ligação.

Se a associação entre os alelos em dois *loci* for não aleatória, diz-se então que os *loci* estão em **desequilíbrio de ligação (LD)**. Nesse caso, um alelo específico no primeiro *locus* está associado a um alelo específico no segundo *locus* mais frequentemente do que o esperado ao acaso. A Figura 18.17 B diagrama um caso de LD completo entre dois *loci*. O alelo *A* sempre está associado ao alelo *B*, enquanto o alelo *a* sempre está associado ao alelo *b*. Não existem cromossomos com haplótipos *Ab* ou *aB*. Nesse caso, as frequências observadas e esperadas não serão as mesmas.

Podemos quantificar o nível de LD entre dois *loci* como a diferença (*D*) entre a frequência observada de um haplótipo e a frequência esperada em relação a uma associação aleatória entre alelos nos dois *loci*. Se ambos os *loci* envolvidos apresentarem apenas dois alelos, então:

$$D = P_{AB} — p_A p_B$$

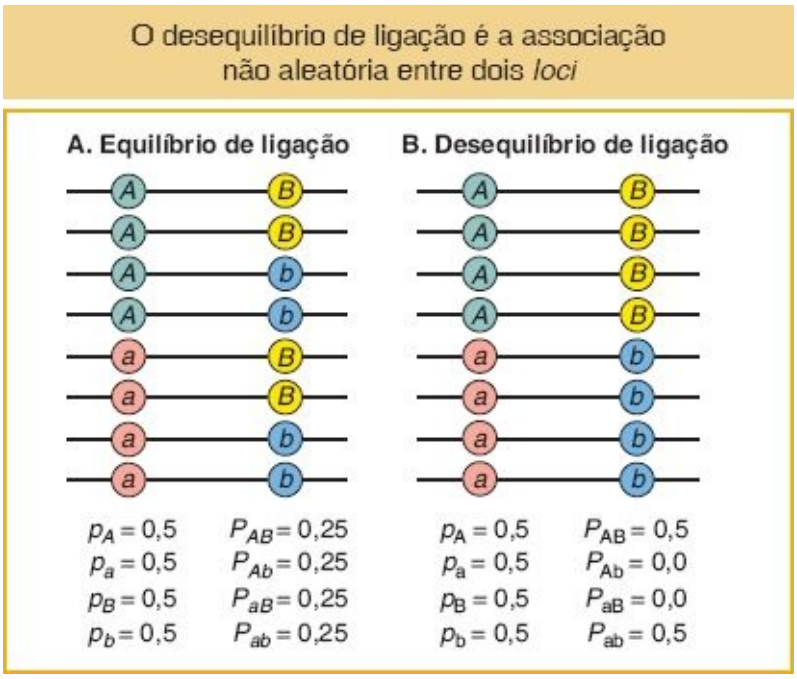


FIGURA 18.17 A. Equilíbrio de ligação. **B.** Desequilíbrio de ligação em relação a dois *loci* (*A* e *B*).

Na Figura 18.17 A, $D = 0$, tendo em vista que não há LD e, na Figura 18.17 B, $D = 0,25$, que é superior a 0, indicando a presença de LD.

Como o LD surge? Sempre que ocorre uma nova mutação em um *locus*, a mutação aparece em um único cromossomo específico e, assim, ela é instantaneamente ligada (ou associada) a alelos específicos em qualquer *loci* vizinho naquele cromossomo. Considere uma população na qual existem apenas dois haplótipos: *AB* e *Ab*. Se uma nova mutação (*a*) surgir no *locus A* em um cromossomo que já possui o alelo *b* no *locus B*, então um novo haplótipo *ab* será formado. Ao longo do tempo, esse novo haplótipo *ab* pode aumentar em frequência na população. Outros cromossomos na população possuiriam os haplótipos *AB* ou *Ab* nesses dois *loci*, mas nenhum cromossomo possuiria *aB*. Portanto, os *loci* estariam em LD. A migração também pode causar LD quando uma subpopulação apresenta apenas o haplótipo *AB* e outra apenas o haplótipo *ab*. Quaisquer migrantes entre as subpopulações dariam origem a LD na subpopulação que recebe os migrantes.

O LD entre dois *loci* declinará ao longo do tempo na medida em que os *crossovers* entre eles tornarem aleatória a relação entre os seus alelos. A taxa de

declínio no LD depende da taxa na qual o *crossing over* ocorre. A frequência de recombinantes (FR) entre os dois *loci* nos gametas que formam a próxima geração (ver Capítulo 4) fornece uma estimativa da taxa de recombinação, a qual, na genética de populações, é simbolizada pela letra minúscula *r*. Se D_0 é o valor do desequilíbrio de ligação entre dois *loci* na geração atual, então o valor na próxima geração (D_1) é fornecido por esta equação:

$$D_1 = D_0(1 — r)$$

Em outras palavras, o desequilíbrio de ligação, conforme medido por *D*, declina a uma taxa de (1 — *r*) por geração. Quando *r* é pequena, *D* declina lentamente ao longo do tempo. Quando *r* está no seu máximo (0,5), então *D* declina em 1/2 a cada geração.

Tendo em vista que o LD decai como uma função do tempo e da fração de recombinação, geneticistas de populações podem utilizar o nível de LD entre uma mutação e os *loci* que a circundam para estimar o tempo em gerações desde que a mutação surgiu pela primeira vez na população. Mutações mais antigas apresentam pouco LD com os *loci* vizinhos, enquanto mutações recentes demonstram um alto nível de LD com os *loci* vizinhos. Se você observar novamente a Figura 18.14, você verá que existe um LD considerável entre *SNP2* no *G6PD* e os SNP vizinhos. *SNP2* codifica a substituição do aminoácido valina por metionina no alelo A^- que confere resistência à malária. Geneticistas de populações utilizaram o LD na *G6PD* para estimar que o alelo A^- surgiu há aproximadamente 10.000 anos. Não se acreditava que a malária fosse prevalente na África até então. Portanto, o A^- surgiu por meio de uma mutação aleatória, mas foi mantido na população em virtude de proporcionar a proteção contra a malária.

> CONCEITO-CHAVE O desequilíbrio de ligação decorre do fato de que novas mutações surgem em um único haplótipo. O desequilíbrio de ligação decairá ao longo do tempo em virtude da recombinação.

Deriva genética e tamanho da população

A lei de Hardy-Weinberg nos informa que as frequências alélicas permanecem as mesmas de uma geração até a próxima em uma *população infinitamente grande*. Entretanto, as reais populações de organismos na natureza são *finitas*, não infinitas. Em populações finitas, as frequências alélicas podem ser alteradas de uma geração para a próxima como resultado do acaso (erro de amostragem) quando os gametas são retirados do *pool* gênico para formar a próxima geração. A alteração nas frequências alélicas entre as gerações em virtude de erro de amostragem é denominada **deriva genética aleatória**, ou apenas deriva, abreviadamente.

Consideraremos um caso simples, porém extremo — uma população composta por um único indivíduo ($N = 1$) heterozigoto (*A/a*) na geração t_0. Possibilitaremos a autofertilização. Nesse caso, o *pool* gênico pode ser descrito como apresentando dois alelos, *A* e *a*, cada um presente a uma frequência de $p = q = 0{,}5$. O tamanho da população permanece o mesmo, $N = 1$, na geração subsequente, t_1. Qual é a probabilidade de que as frequências alélicas sejam alteradas ("derivem") para $p = 1$ e $q = 0$ na geração t_1? Em outras palavras, qual é a probabilidade de que a população se torne fixada em relação ao alelo *A*, de modo que ela seja composta por um único indivíduo homozigoto *A/A*? Tendo em vista que $N = 1$, precisamos coletar apenas dois gametas do *pool* gênico para formar um único indivíduo. A probabilidade de coleta de dois *A* é $p^2 = 0{,}5^2 = 0{,}25$. Portanto, em 25% das ocasiões essa população "derivará" para longe das frequências alélicas originais e se tornará fixada em relação ao alelo *A* após apenas uma geração.

O que ocorre se aumentarmos o tamanho da população para $N = 2$ e o *pool* gênico inicial ainda apresentar $p = q = 0{,}5$? As frequências alélicas serão alteradas para $p = 1$ e $q = 0$ na próxima geração apenas se a população for composta por dois indivíduos *A/A*. Para que isso ocorra, precisamos coletar quatro alelos *A*, cada um com uma probabilidade de $p = 0{,}5$; assim, a probabilidade de que próxima geração apresente $p = 1$ e $q = 0{,}0$ é $p^4 = (0{,}5)^4 = 0{,}0625$, ou apenas pouco mais de 6%. Portanto, uma população com $N = 2$

apresenta menos propensão à deriva para a fixação do alelo *A* do que uma população com $N = 1$. Em termos mais gerais, a probabilidade de que uma população sofra deriva para a fixação do alelo *A* em uma única geração é p^{2N} e, portanto, essa probabilidade se torna progressivamente menor na medida em que o tamanho da população (N) se torna maior. A deriva é uma força mais fraca em populações grandes.

Deriva significa qualquer alteração nas frequências alélicas em virtude de erro de amostragem, não apenas a perda ou a fixação de um alelo. Em uma população de $N = 500$ com dois alelos a uma frequência de $p = q = 0{,}5$, existem 500 cópias de *A* e 500 cópias de *a*. Se a próxima geração apresentar 501 cópias de *A* ($p = 0{,}501$) e 499 cópias de *a* ($q = 0{,}499$), terá ocorrido deriva genética, embora seja um nível de deriva muito modesto. Uma fórmula geral para o cálculo da probabilidade de observação de um número específico de cópias de um alelo na próxima geração, tendo em vista as frequências na geração atual, é apresentada no Quadro 18.4.

Quando a deriva está operando em uma população finita, podemos calcular a probabilidade de diferentes desfechos, mas não podemos prever com precisão o desfecho específico que ocorrerá. O processo é como o de jogar dados. Em qualquer *locus*, a deriva pode continuar de uma geração até a próxima, até que um alelo tenha sido fixado. Além disso, em uma população em particular, a frequência do alelo *A* pode aumentar da geração t_0 para a t_1, mas em seguida diminuir da geração t_1 até a t_2. A deriva não prossegue em um sentido específico em direção à perda ou à fixação de um alelo.

As Figuras 18.18 A e 18.18 B demonstram estudos aleatórios simulados por computador (jogadas de dados) em relação a seis populações de tamanho $N = 10$ e $N = 500$. Cada população inicia com a apresentação de dois alelos a uma frequência de $p = q = 0{,}5$, e em seguida os estudos aleatórios prosseguem durante 30 gerações. Primeiramente, observe a aleatoriedade do processo de uma geração até a próxima. Por exemplo, a frequência de *A* na população ilustrada pela linha amarela na Figura 18.18 A oscila para cima e para baixo de uma geração até a próxima, alcançando uma baixa de $p = 0{,}16$ na t_{16}, mas que em seguida retorna para $p = 0{,}75$ na t_{30}. Em segundo lugar, seja $N = 10$ ou $N = 500$, observe que duas

populações não apresentam exatamente a mesma trajetória. A deriva é um processo aleatório, e provavelmente não observaremos exatamente o mesmo desfecho com populações diferentes ao longo de muitas gerações, exceto quando N é muito pequeno. Em terceiro lugar, observe que quando $N = 10$, as populações se tornaram fixadas (seja $p = 1$ ou $p = 0$) antes da geração 20 em cinco dos seis estudos. Entretanto, quando $N = 500$, as populações retiveram ambos os alelos em todos os seis estudos, até mesmo depois de 30 gerações.

Quadro 18.4 A frequência alélica é alterada sob a deriva.

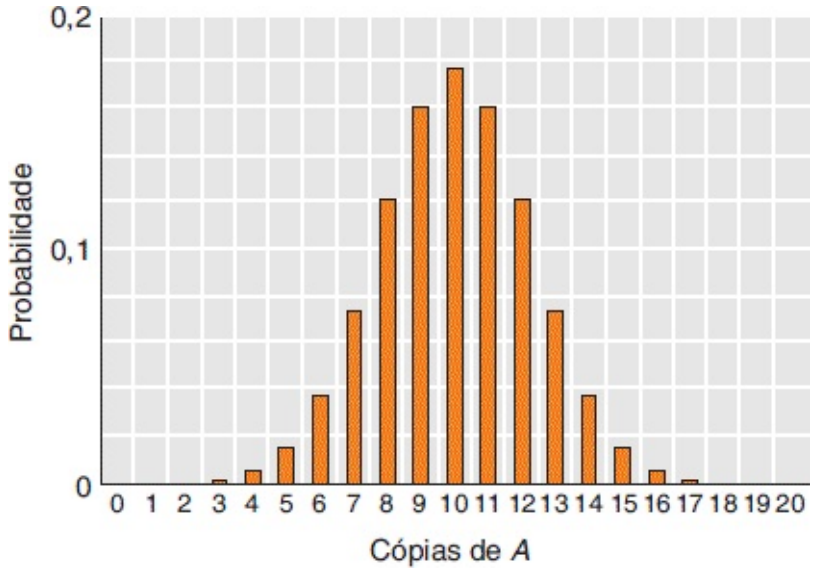
Considere uma população de N indivíduos diploides segregando para dois alelos A e a no *locus* A, com frequências p e q, respectivamente. A população é de cruzamento aleatório e o tamanho da população permanece o mesmo (N) em cada geração. Quando se obtém uma amostra do *pool* gênico para criar a próxima geração, o número exato de cópias do alelo A que são retiradas não pode ser previsto estritamente em virtude de erro de amostragem. Entretanto, a probabilidade de que um número específico de cópias de A seja retirado pode ser calculada com a utilização da fórmula binomial. Deixe que k seja o número específico de cópias do alelo A. A probabilidade de coleta de k cópias é:

$$\text{Prob}(k) = \left(\frac{2N!}{k!\,(2N-k)!} \right) p^k q^{(2N-k)}$$

Se estabelecermos $N = 10$ e $p = q = 0{,}5$, então a probabilidade de coleta de 10 cópias do alelo A é:

$$\text{Prob}(10) = \left(\frac{20!}{10!\,(20-10)!} \right) 0{,}5^{10} 0{,}5^{(20-10)} = 0{,}176$$

Portanto, em apenas 17,6% das ocasiões a próxima geração apresentará a mesma frequência de A e a da geração original. Podemos utilizar essa fórmula para calcular os desfechos em relação a todos os valores possíveis de k e obter uma distribuição de probabilidade, demonstrada na figura a seguir.



Distribuição de probabilidade demonstrando a probabilidade de que números diferentes de *A* estarão presentes após uma geração.

O desfecho único mais provável é a ausência de deriva, com $k = 10$ e uma probabilidade de 0,176. Entretanto, todos os outros desfechos envolvem alguma deriva e, assim, a probabilidade de que a população apresente alguma deriva é de 0,824.

Além do tamanho da população, o destino de um alelo é determinado por sua frequência na população. Especificamente, a probabilidade de que um alelo sofra deriva para fixação em uma geração futura é igual à sua frequência na geração atual. Um alelo que está a uma frequência de 0,5 apresenta uma chance de 50:50 de fixação ou perda da população em uma geração futura. Você pode observar o efeito da frequência alélica sobre o destino de um alelo na Figura 18.18 C. Em relação a dez populações com uma frequência inicial de $p = 0,1$, oito populações apresentaram a perda do alelo *A*, uma a sua fixação, e uma população reteve ambos os alelos após 30 gerações. Isso está muito próximo da expectativa de que *A* chegue à fixação em 10% das ocasiões quando $p = 0,1$.

O fato de a frequência de um alelo ser igual à sua probabilidade de fixação significa que a maior parte das mutações surgidas recentemente será finalmente

perdida em uma população em virtude da deriva. A frequência inicial de uma nova mutação no *pool* gênico é:

$$\frac{1}{2N}$$

Se N for até mesmo modestamente grande, tal como 10.000, a probabilidade de que uma nova mutação finalmente alcance a fixação é, então, extremamente pequena, $1/2N = 1/20.000 = 5 \times 10^{-5}$. A probabilidade de que uma nova mutação finalmente seja perdida da população é:

$$\frac{2N-1}{2N} = 1 - \frac{1}{2N}$$

que está próxima de 1,0 em grandes populações. Ela é de 0,99995 em uma população de 10.000.

A Figura 18.19 A demonstra uma representação gráfica do destino de novas mutações em uma população. O eixo x representa o tempo e o eixo y o número de cópias de um alelo. As linhas pretas demonstram o destino da maior parte das novas mutações. Elas aparecem e logo em seguida são perdidas da população. As linhas coloridas demonstram as poucas novas mutações "de sorte" que se fixaram. A partir da teoria da genética de populações, pode-se demonstrar que o tempo médio necessário para que uma mutação tenha a sorte de ser fixada é de $4N$ gerações. A Figura 18.19 B demonstra uma população que tem 1/2 do tamanho da população na Figura 18.19 A. Portanto, $4N$ gerações correspondem a 1/2 do tempo e as novas mutações de sorte são fixadas mais rapidamente.

Uma consequência importante da deriva é que alelos levemente deletérios podem ser fixados ou alelos vantajosos podem ser perdidos por meio desse processo aleatório. Considere um novo alelo que surge em uma população e que proporciona ao indivíduo que o carrega ter um sistema imune mais forte. Esse indivíduo pode transmitir o alelo vantajoso para a sua descendência, mas a descendência pode morrer antes da reprodução em virtude de um evento aleatório, tal como o fato de ser atingida por raios. Ou, se o indivíduo que carrega o alelo favorável for heterozigoto, ele pode transmitir apenas o alelo menos

favorável para a sua descendência ao acaso.

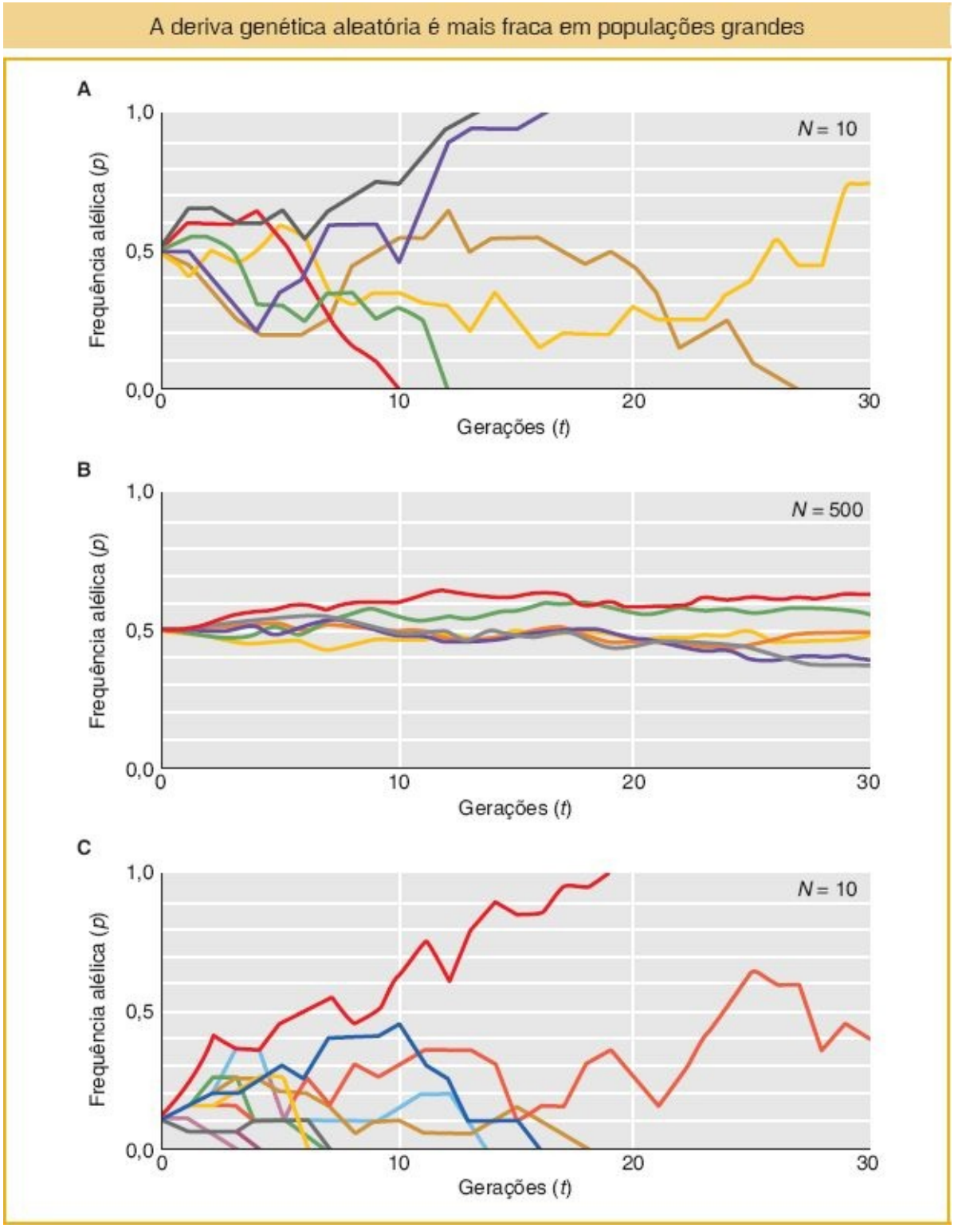


FIGURA 18.18 Simulações em computador de deriva genética aleatória. Cada *linha colorida* representa

uma população simulada ao longo de 30 gerações. **A.** $N = 10$, $p = q = 0{,}5$. **B.** $N = 500$, $p = q = 0{,}5$. C. $N = 10$, $p = q = 0{,}9$.

Ao calcular as probabilidades de diferentes desfechos sob deriva genética, estamos presumindo que os alelos *A* e *a* não conferem diferenças na viabilidade ou no sucesso reprodutivo aos indivíduos que os carregam. Presumimos que indivíduos *A/A*, *A/a* e *a/a* apresentam probabilidade igual de sobreviver e se reproduzir. Nesse caso, *A* e *a* seriam denominados **alelos neutros** (ou variantes) um em relação ao outro. A alteração nas frequências de alelos neutros ao longo do tempo em virtude da deriva é denominada **evolução neutra**. O processo de evolução neutra é o fundamento do **relógio molecular**, a taxa constante de substituição de variantes alélicas de surgimento recente em relação às preexistentes durante longos períodos (Quadro 18.5). A evolução neutra é distinta da evolução darwiniana, na qual os alelos favoráveis aumentam em frequência em virtude de os indivíduos que os carregam deixarem mais descendentes. Discutiremos sobre a evolução darwiniana na próxima seção deste capítulo e no Capítulo 20.

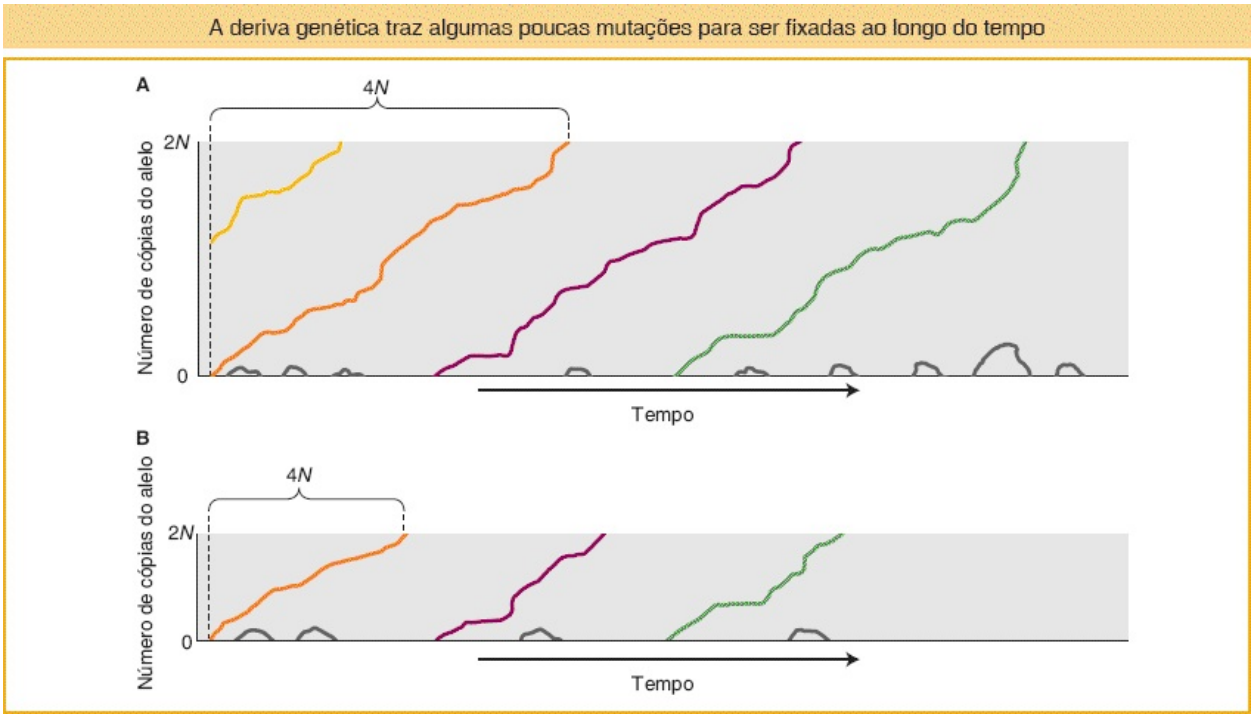


FIGURA 18.19 A. Representação gráfica do aparecimento, da perda e da incorporação final de novas

mutações em uma população ao longo do tempo sob a ação da deriva genética. As linhas de *cor cinza* demonstram o destino da maior parte das novas mutações, as quais aparecem e em seguida são perdidas da população em algumas poucas gerações. As *linhas coloridas* demonstram o destino das poucas mutações "de sorte" que continuam a aumentar em frequência, até alcançarem a fixação. **B.** Uma população que tem 1/2 do tamanho da população na parte **A.** Nessa população, 4*N* gerações correspondem a 1/2 do tempo e as novas mutações de sorte são fixadas mais rapidamente.

Até agora, temos considerado a deriva no contexto de populações que permanecem do mesmo tamanho de uma geração para a próxima. Na realidade, as populações com frequência contraem ou expandem em tamanho ao longo do tempo. Por exemplo, uma nova população de tamanho muito menor pode ser formada subitamente quando um número relativamente pequeno de membros de uma população migra para um novo local e estabelece uma nova população. Os migrantes, ou "fundadores" da nova população podem não carregar todos os alelos presentes na população original, ou podem carregar os mesmos alelos, porém em frequências diferentes. A deriva genética causada pela amostragem aleatória da população original para criar a nova população é conhecida como **efeito fundador**. Um dos muitos eventos fundadores na história humana ocorreu quando pessoas cruzaram a ponte terrestre de Bering da Ásia até as Américas durante a era do gelo, há aproximadamente 15.000 a 30.000 anos. Como resultado, a diversidade genética entre nativos americanos é mais baixa do que entre pessoas em outras regiões do mundo (Figura 18.20).

O tamanho da população também pode ser alterado em um único local. Um período de uma ou diversas gerações consecutivas de contração no tamanho da população é conhecido como um **gargalo** populacional. Os gargalos ocorrem em populações naturais em virtude de flutuações ambientais, tais como uma redução no suprimento alimentar ou um aumento na predação. O lobo-cinzento, o bisão-americano, a águia-de-cabeça-branca, o condor-da-califórnia, o grou-americano e muitas espécies de baleias são alguns exemplos familiares de espécies que sofreram gargalos recentes em virtude da caça pelos seres humanos, ou da invasão do seu *habitat* por parte dos seres humanos. A redução no tamanho da população durante um gargalo aumenta o nível de deriva em uma população. Conforme explicado anteriormente no capítulo, o nível de endocruzamento em

populações também depende do tamanho da população. Portanto, os gargalos também causam um aumento no nível de endocruzamento.

O condor-da-califórnia representa um exemplo extraordinário de um gargalo. Essa espécie chegou a habitar uma ampla área, mas, na década de 1980, ela declinou até uma população reprodutiva de apenas 14 aves em cativeiro. A população atualmente é superior a 400 indivíduos, mas a heterozigosidade média no genoma diminuiu em 8% durante o gargalo inicial. Além disso, um alelo recessivo deletério em relação a um tipo letal de nanismo ocorre a uma frequência de aproximadamente 9% entre os animais sobreviventes, presumivelmente como um resultado da deriva a partir de uma frequência mais baixa na população pré-gargalo. Para tratar desses problemas, biólogos de conservação estabelecem cruzamentos de animais em cativeiro para minimizar o endocruzamento adicional e remover os alelos deletérios da população.

Quadro 18.5 Relógio molecular

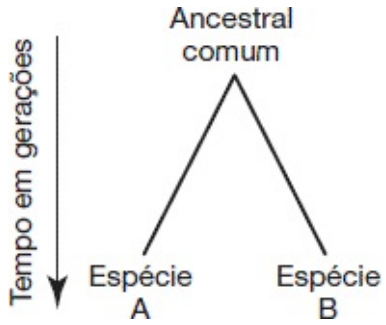
Visto que as espécies divergem ao longo do tempo, suas sequências de DNA se tornam cada vez mais diferentes na medida em que mutações surgem e são fixadas na população. A qual taxa as sequências divergem? Para responder a essa questão, considere uma população na geração t_0. O número de mutações que aparecerá na geração t_1 é o produto do número de cópias da sequência no *pool* gênico ($2N$) vezes a taxa na qual elas sofrem mutação (μ); ou seja, $2N\mu$. Se uma mutação for neutra, então a probabilidade de que ela derive até a fixação é de $1/2N$. Assim, a cada geração, $2N\mu$ novas mutações entram no *pool* gênico, e $1/2N$ delas será fixada. O produto desses dois números é a taxa (k) na qual as sequências evoluem:

$$k = 2N\mu \times \frac{1}{2N} = \mu$$

O valor k é denominado taxa de substituição, e é igual à taxa de mutação em relação às mutações neutras. Se a taxa de mutação permanecer constante ao longo do tempo, a taxa de substituição então “baterá” regularmente como um

relógio, o relógio molecular.

Considere duas espécies, A e B, e seu ancestral comum. Definiremos *d* (divergência) como o número de substituições neutras em sítios de nucleotídios na sequência de DNA de um gene que ocorreram desde a divergência de A e B de seu ancestral.



O valor esperado em relação a *d* será o produto da taxa (*k*) na qual as substituições ocorrem e duas vezes o tempo em gerações (2*t*) durante as quais a substituição foi acumulada. O 2 é necessário em virtude de existirem duas linhagens que causam o distanciamento do ancestral comum. Portanto, temos:

$$d = 2tk$$

Essa equação pode se reescrita como:

$$t = \frac{d}{2k}$$

demonstrando como podemos calcular o tempo em gerações desde a divergência de duas espécies quando conhecemos *d* e *k*. A taxa de mutação de SNP por geração (μ) é conhecida em relação a muitos grupos de organismos (ver Tabela 18.5), e é a mesma que a taxa de substituição (*k*) em relação às mutações neutras. Pode-se sequenciar um ou mais genes de duas espécies e determinar a proporção de sítios de nucleotídios silenciosos (neutros) nos quais eles diferem e utilizar essa proporção como uma estimativa em relação a *d*. Portanto, pode-se calcular o tempo desde que duas sequências (duas espécies) divergiram com a utilização do relógio molecular. Entre seres

humanos e chimpanzés, existe aproximadamente 0,018 diferença de bases em sítios sinônimos nas sequências codificadoras. A taxa de mutação de SNP em relação aos seres humanos é de 3×10^{-8}, e o tempo de geração é de aproximadamente 20 anos. Com a utilização desses valores e da equação anterior, o tempo de divergência estimado em relação aos seres humanos e aos chimpanzés é de aproximadamente 6,0 milhões de anos. Esses cálculos presumem que as substituições são neutras e que a taxa de substituição tem sido constante ao longo do tempo.

O Quadro 18.6 discute sobre o gargalo bem-caracterizado que ocorreu durante a domesticação de espécies de cultivo. Esse gargalo explica o motivo de nossas plantas de cultivo apresentarem muito menos diversidade genética do que seus ancestrais selvagens.

CONCEITO-CHAVE O tamanho da população é um fator-chave que afeta a variação genética nas populações. A deriva genética exerce maior força sobre populações pequenas do que grandes. A probabilidade de que um alelo se torne fixado (ou seja perdido) em uma população por meio de deriva é uma função de sua frequência na população e do tamanho da população. A maior parte das novas mutações neutras é perdida das populações em decorrência de deriva.

Seleção

Até agora, consideramos como os novos alelos entram em uma população por meio de mutação e migração e como esses alelos podem se fixar (ou se perder) em uma população por meio da deriva aleatória. Mas a mutação, a migração e a deriva não conseguem explicar o motivo de os organismos aparentarem estar tão bem-adaptados aos seus ambientes. Elas não conseguem explicar as **adaptações**, as características da forma ou da fisiologia de um organismo que possibilitam que ele lide melhor com as condições ambientais sob as quais ele vive. Para explicar

a origem das adaptações, em 1859, Charles Darwin, em seu histórico livro *A Origem das Espécies*, propôs que as adaptações surgem por meio da ação de outro processo, que ele denominou "seleção natural". Nesta seção, exploraremos o papel da seleção natural na modulação da variação genética nas populações. Posteriormente, no Capítulo 20, consideraremos os efeitos da seleção natural sobre a evolução de genes e de traços ao longo de períodos prolongados.

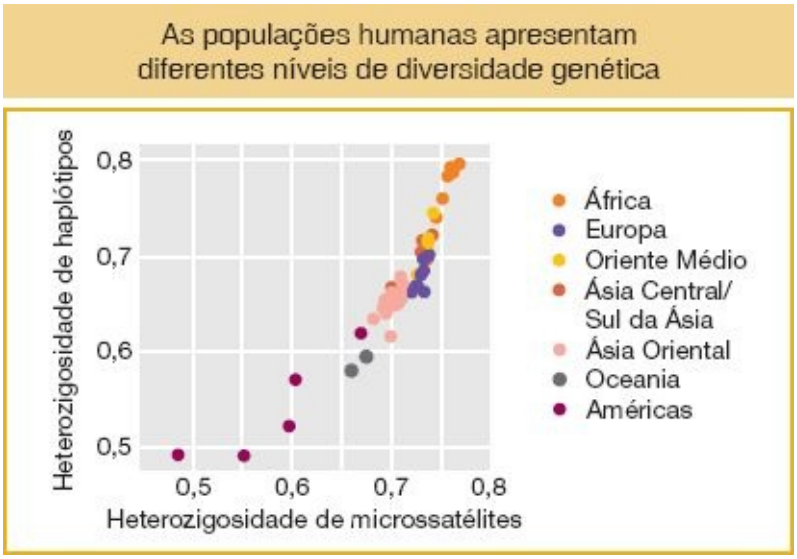


FIGURA 18.20 O gráfico da heterozigosidade de haplótipos *versus* heterozigosidade de microssatélites demonstra a diversidade genética em relação a diferentes grupos geográficos de seres humanos. A diversidade genética é mais baixa para os nativos americanos em virtude do efeito fundador. (Dados de D. F. Conrad *et al.*, *Nat. Genet.* 38, 2006, 1251-1260.)

Quadro 18.6 Gargalo da domesticação.

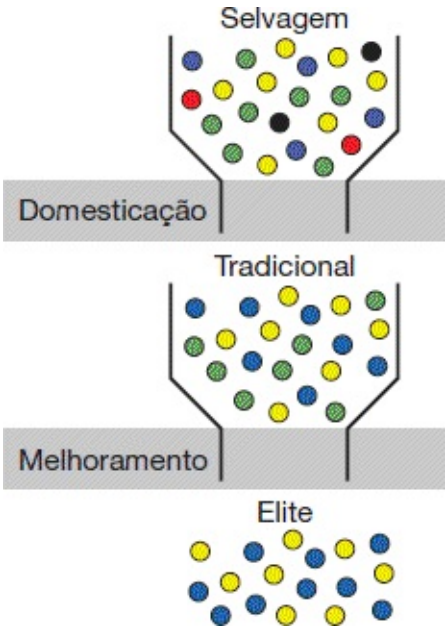
Antes de 10.000 anos atrás, nossos ancestrais em todo o mundo proviam a si mesmos com a caça de animais selvagens e a coleta de plantas alimentícias selvagens. Aproximadamente naquela época, as sociedades humanas começaram a desenvolver a agricultura. Pessoas coletavam plantas e animais selvagens locais, cultivavam plantações e domesticavam os animais. Alguns dos principais cultivos que foram domesticados nessa ocasião incluem o trigo no Oriente Médio, o arroz na Ásia, o sorgo na África e o milho no México.

Quando os primeiros fazendeiros coletaram sementes selvagens para iniciar

a domesticação, eles coletaram uma amostra do *pool* gênico selvagem. Essa amostra apresentava apenas um subconjunto da variação genética observada na vida selvagem. As populações domesticadas passaram por um gargalo. Consequentemente, as plantas de cultivo e os animais domesticados tipicamente apresentam menos variação genética do que os seus genitores selvagens.

O cultivo científico moderno de plantas que objetivava a melhora das plantações criou um segundo gargalo. Ao amostrar o *pool* gênico das variedades de cultivo tradicionais, agricultores modernos criaram variedades de elite, com traços de valor comercial, tais como alta produção e adequabilidade à coleta e ao processamento mecânicos. Consequentemente, as variedades de elite ou modernas apresentam ainda menos variação genética do que as variedades tradicionais.

A perda da variação genética resultante dos gargalos da domesticação e do melhoramento pode impor uma ameaça. Tendo em vista que existem menos alelos por *locus*, os cultivos apresentam um repertório menor de alelos nos genes de resistência a doenças e possivelmente maior suscetibilidade a patógenos emergentes. Para reduzir essa vulnerabilidade, cultivadores realizam cruzamentos entre as variedades modernas e os parentes selvagens (ou as variedades tradicionais) para reintroduzir alelos criticamente importantes nos cultivos modernos.



Gargalos da domesticação e do melhoramento de cultivos. Os *pontos coloridos* representam alelos diferentes. (M. Yamasaki *et al*., *Plant Cell* 17, 2005, 2859-2872.)

Definiremos a **seleção natural** como o processo por meio do qual os indivíduos com determinadas características hereditárias apresentam maior probabilidade de sobreviver e se reproduzir do que outros indivíduos com ausência dessas características. Conforme resumido por Darwin, o processo funciona assim. A cada geração, são produzidos mais descendentes do que os que conseguem sobreviver e se reproduzir no ambiente. A natureza apresenta um mecanismo (mutação) para produzir novas formas ou variantes hereditárias. Indivíduos com variantes particulares de algumas características apresentam maior probabilidade de sobreviver e se reproduzir. Indivíduos com características que intensificam a sua capacidade de sobreviver e se reproduzir transmitirão essas características à sua descendência. Ao longo do tempo, essas características aumentarão em frequência na população. Portanto, as populações serão alteradas ao longo do tempo (*evoluirão*) na medida em que o ambiente (*a natureza*) favorecer (*selecionar*) as características que intensificam a capacidade

de sobreviver e se reproduzir. Essa é a teoria da evolução de Darwin por meio da seleção natural.

A evolução darwiniana com frequência é descrita com a utilização da frase "*sobrevivência do mais apto*". Essa frase pode ser enganosa. Um indivíduo que é fisicamente forte, resistente a doenças e vive uma vida longa, mas que não apresenta descendência, não é adaptado no sentido darwiniano. A **adaptabilidade darwiniana** refere-se à capacidade de sobreviver e de *se reproduzir*. Ela considera tanto a viabilidade quanto a fecundidade. Uma medida da adaptabilidade darwiniana é simplesmente o tamanho da descendência que um indivíduo apresenta. Essa medida é denominada **adaptabilidade absoluta**, e a simbolizaremos com um *W* maiúsculo. Em relação a um indivíduo sem descendência, *W* é igual a 0, em relação a um indivíduo com um descendente, *W* é igual a 1, em relação a um indivíduo com dois descendentes, *W* é igual a 2, e assim por diante. *W* também é o número de alelos em um *locus* com que um indivíduo contribui para o *pool* gênico.

A adaptabilidade absoluta mistura o tamanho da população e diferenças no sucesso reprodutivo entre os indivíduos. Geneticistas de populações estão primariamente interessados no último e, assim, utilizam uma medida denominada **adaptabilidade relativa** (simbolizada por um *w* minúsculo), que é a adaptabilidade de um indivíduo em relação àquela de algum outro indivíduo, normalmente o indivíduo mais adaptado na população. Se o indivíduo X apresenta dois descendentes e o indivíduo mais adaptado, Y, apresenta 10 descendentes, a adaptabilidade relativa de X é $w = 2/10 = 0{,}2$. A adaptabilidade relativa de Y é $w = 10/10 = 1$. Para cada 10 alelos com os quais Y contribui para a próxima geração, X contribuirá com 2.

O conceito de adaptabilidade se aplica aos genótipos, bem como aos indivíduos. A adaptabilidade absoluta em relação ao genótipo *A/A* ($W_{A/A}$) é o número médio de descendentes deixados pelos indivíduos com aquele genótipo. Se conhecermos a adaptabilidade absoluta em relação a todos os genótipos em um *locus*, podemos então calcular a adaptabilidade relativa de cada um dos genótipos.

Agora veremos como as frequências alélicas podem ser alteradas ao longo do

tempo quando genótipos diferentes apresentam adaptabilidade diferente; ou seja, quando a seleção natural está atuando. A seguir estão a adaptabilidade e as frequências genotípicas em relação a três genótipos no *locus A* em uma população. Nesse caso, *A* é um alelo dominante favorecido, tendo em vista que a adaptabilidade dos indivíduos *A/A* e *A/a* é a mesma e superior à dos indivíduos *a/a*. Estamos presumindo que essa população segue a lei de Hardy-Weinberg, com $p = 0,1$ e $q = 0,9$.

	A/A	*A/a*	*a/a*
Número médio de descendentes (*W*)	10	10	5
Adaptabilidade relativa (*w*)	1,0	1,0	0,5
Frequência genotípica	0,01	0,18	0,81

A contribuição relativa de cada genótipo para o *pool* gênico é determinada pelo produto de sua adaptabilidade com sua frequência. Quanto mais adaptado e mais alta a frequência de um genótipo, mais ele contribui.

Genótipo	*A/A*	*A/a*	*a/a*	Soma
Contribuição relativa	$1 \times 0,01 = 0,01$	$1 \times 0,18 = 0,18$	$0,5 \times 0,81 = 0,405$	0,595

As contribuições relativas não somam 1, então precisamos redimensioná-las ao

dividir cada uma pela soma das três (0,595) para obter as frequências genotípicas esperadas que contribuem para o *pool* gênico.

Genótipo	*A/A*	*A/a*	*a/a*	Soma
Frequências genotípicas	0,02	0,30	0,68	1,0

Com a utilização dessas frequências genotípicas esperadas e da lei de Hardy-Weinberg, podemos calcular as frequências alélicas na próxima geração:

$$p' = 0{,}02 + (\tfrac{1}{2} \times 0{,}3) = 0{,}17$$

e:

$$q' = 0{,}68 + (\tfrac{1}{2} \times 0{,}3) = 0{,}83$$

A diferença entre p' e q' ($\Delta p = p' - p$) é 0,17 — 0,1 = 0,07, então concluímos que o alelo *A* aumentou 7% em uma geração em virtude da seleção natural. O Quadro 18.7 apresenta as equações-padrão para o cálculo das alterações nas frequências alélicas ao longo do tempo em virtude da seleção natural.

Passaremos por esse processo recursivamente, utilizando as frequências alélicas da primeira geração para calcular aquelas na segunda geração, utilizando em seguida aquelas da segunda para calcular as da terceira e assim por diante. Se em seguida inserirmos em um gráfico p pelo tempo medido em número de gerações (t), teremos um quadro do tempo no qual as frequências alélicas são alteradas sob a força da seleção natural. A Figura 18.21 demonstra um referido gráfico em relação a ambos um alelo dominante favorecido e um recessivo favorecido. O alelo dominante no início aumenta rapidamente, mas em seguida alcança um platô e se aproxima apenas lentamente da fixação. Após o alelo dominante favorecido alcançar uma alta frequência, o alelo recessivo não favorecido ocorre principalmente em heterozigotos e raramente em homozigotos

com adaptabilidade reduzida e, assim, a seleção é ineficaz para removê-lo da população. O recessivo favorecido se comporta da maneira oposta — primeiramente ele aumenta lentamente em frequência, tendo em vista que os homozigotos *a/a* com adaptabilidade intensificada são raros, mas posteriormente prossegue mais rapidamente até a fixação. Tendo em vista que a classe heterozigota apresenta adaptabilidade reduzida, o alelo dominante não favorecido finalmente pode ser removido da população.

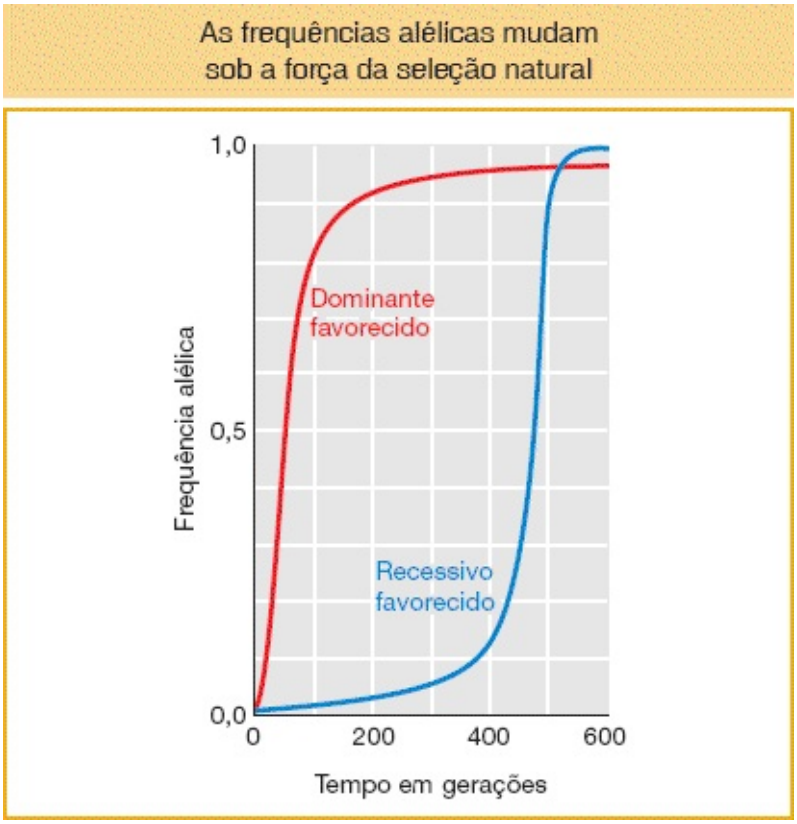


FIGURA 18.21 Alteração na frequência alélica de um alelo dominante favorecido (*vermelho*) e um alelo recessivo favorecido (*azul*), direcionada pela seleção natural ao longo de 600 gerações.

Quadro 18.7 Efeito da seleção sobre as frequências alélicas.

A seleção causa alterações nas frequências alélicas entre gerações, tendo em vista que alguns genótipos contribuem com mais alelos para o *pool* gênico do que outros. Descreveremos um conjunto de equações para prever as frequências gênicas na próxima geração quando a seleção está operando. As

frequências genotípicas e a adaptabilidade absoluta são simbolizadas como segue:

genótipo	*A/A*	*A/a*	*a/a*
frequência	p^2	$2pq$	q^2
adaptabilidade absoluta	$W_{A/A}$	$W_{A/a}$	$W_{a/a}$

O número médio de alelos com que indivíduos de um determinado genótipo contribuem é a frequência do genótipo vezes a adaptabilidade absoluta. Se *N* é o tamanho da população, o número total de alelos com que todos os indivíduos de determinado genótipo contribuem é *N* multiplicado pelo número médio de alelos com que indivíduos de determinado genótipo contribuem:

número médio	$p^2W_{A/A}$	$2pqW_{A/a}$	$q^2W_{a/a}$
número total	$N(p^2)W_{A/A}$	$N(2pq)W_{A/a}$	$N(q^2)W_{a/a}$

Portanto, o *pool* gênico apresentará:

número de alelos $A = N(p^2)W_{A/A} + \frac{1}{2}[N(2pq)W_{A/a}]$
número de alelos a = $N(p^2)W_{a/a} + \frac{1}{2}[N(2pq)W_{A/a}]$

A adaptabilidade média da população é:

$$\overline{W} = p^2 W_{A/A} + 2pqW_{A/a} + q^2 W_{a/a}$$

que é o número médio de alelos com que um indivíduo contribui para o *pool* gênico. $N\overline{W}$ é o número total de alelos no *pool* gênico.

Agora podemos calcular a proporção de alelos *A* no *pool* gênico na próxima geração como:

$$p' = \frac{Np^2 W_{A/A} + NpqW_{A/a}}{N\overline{W}}$$

Essa equação é reduzida para:

$$p' = p\frac{pW_{A/A} + qW_{A/a}}{\overline{W}}$$

Observe a expressão $pW_{A/A} + qW_{A/a}$. Essa é denominada adaptabilidade alélica ou adaptabilidade média de alelos *A* (W_A):

$$W_A = pW_{A/A} + qW_{A/a}$$

A partir da lei de Hardy-Weinberg, sabemos que uma proporção *p* de todos os alelos *A* está presente em homozigotos com outro *A*, caso em que eles apresentam uma adaptabilidade de $W_{A/A}$, enquanto uma proporção *q* de todos os alelos *A* está presente em heterozigotos com *a* e eles apresentam uma adaptabilidade de $W_{A/a}$. Substituindo W_A na equação anterior, obtemos:

$$p' = p\frac{W_A}{\overline{W}}$$

Essa equação pode ser utilizada para calcular a frequência de *A* na próxima geração e pode ser utilizada recursivamente para seguir a alteração em *p* ao longo do tempo.

Embora tenhamos derivado essas fórmulas com a utilização da adaptabilidade absoluta, em geral não estamos interessados no tamanho da população; assim, utilizamos formas dessas equações com a adaptabilidade relativa:

$\bar{w} = p^2 w_{A/A} + 2pq w_{A/a} + q^2 w_{a/a}$

$w_A = p w_{A/A} + q w_{A/a}$

$p' = p\frac{w_A}{\bar{w}}$

Finalmente, podemos expressar a alteração na frequência alélica entre as gerações como:

$$\Delta p = p' - p = p\frac{w_A}{\bar{w}} - p$$
$$= \frac{p(w_A - \bar{w})}{\bar{w}}$$

Mas $\bar{w}$, a adaptabilidade relativa média da população, é a média de w_A e w_a, que são as adaptabilidades alélicas de *A* e *a*, respectivamente:

$$\bar{w} = p w_A + q w_a$$

Substituindo essa expressão por *w* na fórmula em relação a Δp e relembrando que $q = 1 - p$, obtemos:

$$p = \frac{pq(w_A - w_a)}{\bar{w}}$$

Tipos de seleção

A seleção natural pode operar de diversos modos. A **seleção direcional**, que temos discutido, movimenta a frequência de um alelo em uma direção até que ele alcance a fixação ou seja perdido. A seleção direcional pode ser *positiva* ou *purificadora*. A **seleção positiva** atua para trazer uma nova mutação ou um novo alelo favorável até uma frequência mais alta. Esse tipo de seleção está em atuação quando novas adaptações evoluem. Uma *varredura seletiva* ocorre quando um alelo favorável alcança a fixação. A seleção direcional também pode atuar para remover mutações deletérias da população. Esse tipo de seleção é denominado **seleção purificadora**, e evita que as características adaptativas

existentes sejam degradadas ou perdidas. A seleção nem sempre prossegue de modo direcional até a perda ou a fixação de um alelo. Se a classe heterozigota apresenta adaptabilidade mais alta do que qualquer uma das classes homozigotas, a seleção natural então favorecerá a manutenção de ambos os alelos na população. Nesse caso, o *locus* está sob **seleção balanceadora**, e a seleção natural irá movimentar a população até um ponto de equilíbrio, no qual ambos os alelos são mantidos na população (ver Capítulo 20).

Cada um dos diferentes tipos de seleção deixa uma assinatura distinta na sequência de DNA próxima do *locus*-alvo em uma população. Por exemplo, a seleção positiva pode ser detectada em sequências de DNA por meio de seus efeitos sobre a diversidade genética e o equilíbrio de ligação. A Figura 18.22 demonstra haplótipos esquemáticos antes e depois de um episódio de seleção positiva. No painel que demonstra os haplótipos antes da seleção, a região entre chaves apresenta muitos polimorfismos e múltiplos haplótipos. Entretanto, após a seleção, existe apenas um único haplótipo nessa região e, portanto, nenhum polimorfismo. Quando a seleção é aplicada ao sítio-alvo (demonstrado em *vermelho*), todos os sítios-alvo e vizinhos podem ser varridos até a fixação antes que a recombinação fragmente o haplótipo no qual a mutação favorável ocorreu pela primeira vez. O resultado é uma diversidade mais baixa e um LD mais alto próximo do alvo. Na medida em que a distância do alvo aumenta, existem mais oportunidades para a recombinação e, assim, a diversidade retrocede gradualmente.

A Figura 18.23 demonstra o padrão de diversidade na região que circunda o gene *SLC24A5* em seres humanos. Esse gene influencia a deposição da melanina na pele. Quando pessoas migraram da África para a Europa, uma varredura seletiva no *SLC24A5* causou a perda de toda a diversidade nesse *locus*. Consequentemente, existe um alelo único e um haplótipo único nesse *locus* na Europa. O alelo único que foi selecionado na Europa produz a cor da pele mais clara. Com o distanciamento do gene em qualquer direção, o número de haplótipos aumenta nas populações europeias, tendo em vista que a recombinação perturbou o desequilíbrio de ligação entre o *SLC24A5* e os sítios mais distantes. A pele clara pode ser adaptativa em latitudes ao norte. As pessoas são capazes de

sintetizar vitamina D, mas para tanto, elas precisam absorver a radiação UV pela pele. Nas latitudes equatoriais, as pessoas são expostas a altos níveis de luz UV e conseguem sintetizar a vitamina D até mesmo com a pele fortemente pigmentada. Mais distante do equador, as pessoas são expostas a menos luz UV, e a cor da pele mais clara pode facilitar a síntese de vitamina D nessas latitudes.

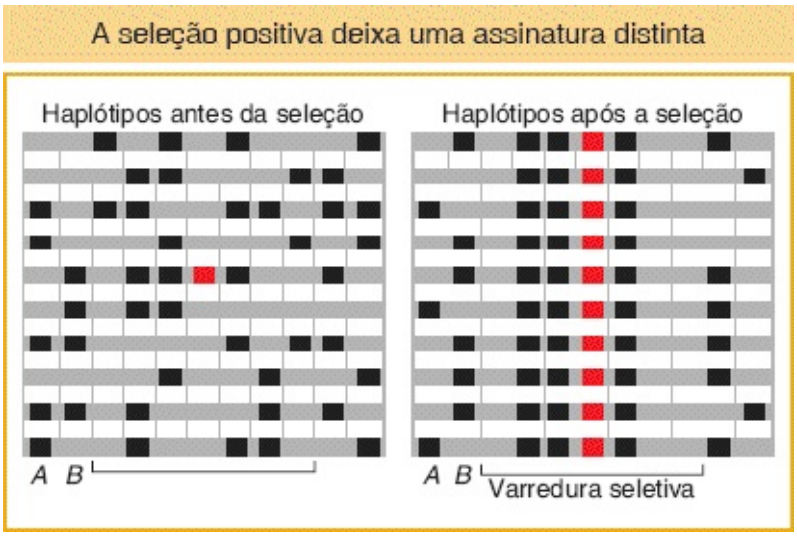


FIGURA 18.22 Esquema de haplótipos observados em uma população antes e depois de um alelo favorecido (*vermelho*) ser varrido até a fixação. Existem 11 *loci* no total. Existem dois alelos (*vermelho* e *cinza*) no *locus* que era o alvo da seleção. Existem dois alelos (*preto* e *cinza*) em cada *locus* ligado ao *locus*-alvo. Após a seleção, o sítio-alvo e alguns vizinhos foram varridos até a fixação.

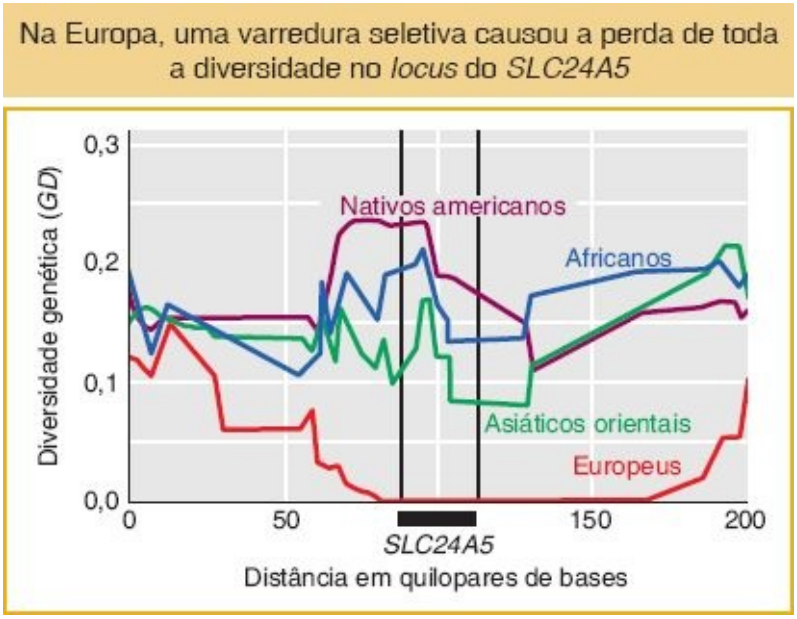


FIGURA 18.23 Diversidade genética em grupos continentais humanos ao longo de um segmento de 2 milhões de pb do cromossomo 15 humano que circunda o gene *SLC24A5*. (Dados de Human Diversity Genome Project, www.hgdp.uchicago.edu.)

A Tabela 18.6 lista alguns dos genes que demonstram evidência de seleção natural em seres humanos modernos. Esses genes estão situados em algumas poucas categorias básicas. Um grupo fortalece a resistência a patógenos. Os genes *G6PD*, FY^{nulo} e *Hb* (*hemoglobina B*, o gene da anemia falciforme), todos, ajudam a adaptar os seres humanos à ameaça da malária. A Figura 18.11 B demonstra que a frequência de FY^{nulo} é mais alta na África central. A África central também apresenta a mais alta prevalência de malária, sugerindo que a seleção levou o FY^{nulo} até a sua mais alta frequência na região na qual a pressão da seleção é maior. Recentemente, geneticistas descobriram o gene *CCR5* (*receptor de quimiocina 5*), que apresenta um alelo (*CCR5*-Δ*32*) que proporciona resistência à AIDS. Esse alelo atualmente é um alvo da seleção natural. Enquanto existirem patógenos, a seleção natural continuará a operar nas populações humanas.

Outro grupo de genes selecionados na Tabela 18.6 adapta as pessoas às dietas regionais. Antes de 10.000 atrás, todos os seres humanos eram caçadores-coletores. Mais recentemente, a maior parte dos seres humanos realizou a transferência para os alimentos agrícolas, mas existem diferenças regionais nas dietas. No norte da Europa e em partes da África, os laticínios são uma parte substancial da dieta. Na maior parte das populações, a enzima lactase para a digestão do açúcar do leite (lactose) é expressa durante a infância, mas deixa de ser sintetizada em adultos. Entretanto, em partes da Europa e da África nas quais os adultos bebem leite, alelos especiais do gene da *lactase*, que continuam a expressar a enzima lactase durante a fase adulta, aumentaram em frequência em virtude da seleção natural. Finalmente, a Tabela 18.6 inclui alguns genes para adaptação fisiológica ao clima. Entre esses estão os genes da pigmentação da pele, tais como o *SLC24A5*, discutido anteriormente.

Tabela 18.6 Alguns genes que demonstram evidência de seleção natural em populações humanas específicas.

Gene	Traço presumido	População
EDA2R (receptor A2 da ectodisplasina)	Calvície-padrão masculina	Europeus
EDAR (receptor A da ectodisplasina)	Morfologia dos cabelos	Asiáticos orientais
FY^{nulo} (antígeno *Duffy*)	Resistência à malária	Africanos
G6PD (glicose-6-fosfato desidrogenase)	Resistência à malária	Africanos
Hb (hemoglobina B)	Resistência à malária	Africanos
KITLG (ligante KIT)	Pigmentação da pele	Asiáticos orientais e europeus
LARGE (glicosiltransferase)	Resistência à febre de Lassa	Africanos
LCT (lactase)	Persistência da lactase; capacidade de digerir o açúcar do leite quando adulto	Africanos, europeus
LPR (receptor de leptina)	Processamento de gorduras alimentares	Asiáticos orientais

MC1R (receptor de melanocortina)	Pigmentação dos cabelos/pelos e da pele	Asiáticos orientais
MHC (complexo principal de histocompatibilidade)	Resistência a doenças infecciosas	Múltiplas populações
OCA2 (albinismo oculocutâneo)	Pigmentação da pele e cor dos olhos	Europeus
PPARD (receptor delta ativado por proliferador de peroxissomo)	Processamento de gorduras alimentares	Europeus
SI (sacarase-isomaltase)	Metabolismo da sacarose	Asiáticos orientais
SLC24A5 (família transportadora de soluto 24)	Pigmentação da pele	Europeus e asiáticos ocidentais
TYRP1 (proteína 1 relacionada à tirosinase)	Pigmentação da pele	Europeus

Fonte: P. C. Sabeti *et al.*, *Science* 312, 2006, 1614-1620; P. C. Saberi *et al. Nature* 449, 2007, 913-919; B. F. Voight *et al. PLoS Biology* 4, 2006, 446-458; J. K. Pickrell *et al.*, *Genome Research* 19, 2009, 826-837.

Enquanto a seleção direcional causa uma perda de variação genética na região que circunda o *locus*-alvo, a seleção balanceadora consegue prevenir a perda de

diversidade por meio de deriva genética aleatória, levando a regiões de diversidade genética incomumente alta no genoma. Uma região de alta diversidade genética circunda o complexo principal de histocompatibilidade (MHC) no cromossomo 6. A Figura 18.24 demonstra um pico distinto no número de SNP no MHC. Esse complexo inclui os genes do antígeno leucocitário humano (HLA), que estão envolvidos no reconhecimento de (e na resposta a) patógenos por parte do sistema imune. A seleção balanceadora é uma hipótese proposta para explicar a alta diversidade observada no MHC. Tendo em vista que heterozigotos apresentam dois alelos, eles podem ser resistentes a um repertório maior de tipos de patógenos, proporcionando aos heterozigotos uma vantagem adaptativa.

Finalmente, a seleção pode ser imposta por outro agente além da natureza. Os seres humanos impuseram a seleção no processo de domesticação e melhoria de cultivos de plantas e criação de animais. Esse tipo de seleção é denominado **seleção artificial**. Nesse caso, os indivíduos com traços que os seres humanos preferem contribuem com mais alelos para o *pool* gênico do que os indivíduos com os traços não favorecidos. Ao longo do tempo, os alelos que conferem os traços favorecidos aumentam em frequência na população. As muitas raças de cães e vacas leiteiras e de variedades de vegetais de jardinagem e cultivos de cereais são, todas, produtos de seleção artificial.

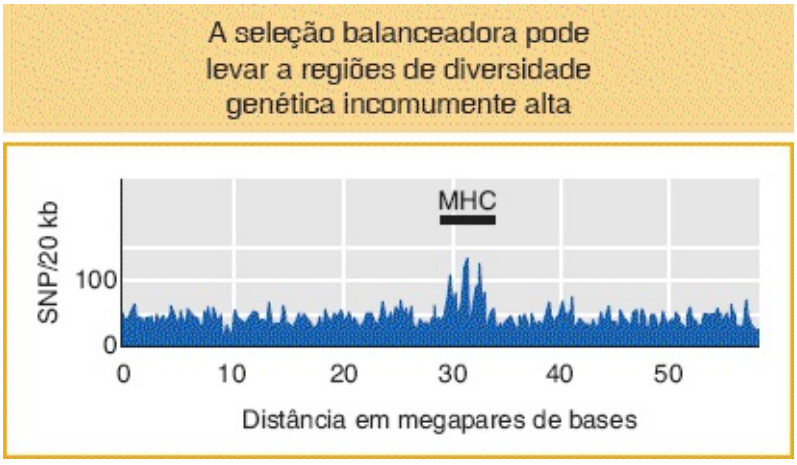


FIGURA 18.24 Número de sítios segregantes (S) ou SNP em janelas de 20 quilopares de bases ao longo do braço curto do cromossomo 6 humano. Existe um pico de alta diversidade no *locus* do MHC. (Dados de International HapMap Project, www.hapmap.org.)

CONCEITO-CHAVE A seleção natural é uma força que pode direcionar alelos favoráveis em um *locus* até a fixação, ou manter múltiplos alelos em um *locus* em uma população. A seleção deixa uma assinatura no genoma, no tipo de padrão de diversidade genética que circunda o alvo da seleção. Geneticistas de populações identificaram um número de genes que têm sido alvo de seleção em seres humanos.

Equilíbrio entre mutação e deriva

Consideramos as forças que regulam a variação nas populações individualmente. Agora consideraremos os efeitos opostos da mutação e da deriva, a primeira adicionando variação e a última removendo-a das populações. Quando essas duas forças estão em equilíbrio, uma população pode alcançar um equilíbrio no qual a perda e o ganho de variação são iguais. Utilizaremos a heterozigosidade (*H*) como uma medida de variação. Relembre que *H* estará próximo de 0 quando uma população estiver próxima da fixação em relação a um único alelo (variação baixa), e que *H* se aproximará de 1 quando houver muitos alelos de frequência igual (variação alta).

Utilizaremos *H* com acento circunflexo, $\hat{H}$, como símbolo do valor de equilíbrio de *H*. Para encontrar , iniciamos com duas equações matemáticas: uma equação que relaciona a alteração em *H* ao tamanho da população (deriva) e outra equação que relaciona a alteração em *H* à taxa de mutação. Em seguida, podemos estabelecer essas equações como iguais entre si e calcular $\hat{H}$.

Primeiramente, precisamos de uma equação para o declínio na variação (*H*) entre as gerações como uma função do tamanho da população (deriva). Desenvolvemos uma referida equação no Quadro 18.3 quando discutimos sobre o endocruzamento:

$$H' = \left(1 - \frac{1}{2N}\right)H$$

Essa equação se aplica aos efeitos da deriva, bem como àqueles do

endocruzamento. A partir dessa equação, segue-se que a alteração em *H* entre as gerações em virtude da deriva é:

$$H = H - H' = \frac{1}{2N} H$$

Em segundo lugar, precisamos de uma equação para o aumento na variação, conforme medido por *H*, entre gerações em virtude da mutação. Qualquer mutação nova aumentará a heterozigosidade a uma taxa proporcional à frequência de homozigotos na população (1 — *H*) vezes a taxa na qual a mutação os converte em heterozigotos (2μ). (O 2 é necessário em virtude da existência de dois alelos que podem sofrer mutação em um diploide.) Portanto, a alteração em *H* entre as gerações em virtude de mutação é:

$$\Delta H = 2\mu(1 — \hat{H})$$

Quando a população alcançar um equilíbrio, a perda da heterozigosidade por meio de deriva será igual ao ganho por mutação. Portanto, temos:

$$\frac{1}{2N} \hat{H} = 2\mu(1 - \hat{H})$$

que pode ser reescrita como:

$$\hat{H} = \frac{4N\mu}{4N\mu + 1}$$

Essa equação fornece o valor de equilíbrio de $\hat{H}$ quando a perda pela deriva e o ganho por mutação estão equilibrados. Essa equação é aplicável apenas à variação neutra; ou seja, estamos presumindo que a seleção não está atuando. Também estamos presumindo que cada nova mutação produz um alelo único.

Expressões tais como essa são úteis quando temos estimativas em relação a duas das variáveis e gostaríamos de conhecer a terceira. Por exemplo, a diversidade de nucleotídios (*H* no nível de nucleotídio) em relação às sequências não codificadoras, que são amplamente neutras, é de aproximadamente 0,0013 em

seres humanos, e μ em seres humanos é 3×10^{-8} (ver Tabela 18.5). A utilização desses valores e a solução da equação anterior em relação a N produzem uma estimativa do tamanho da população humana de 10.498 indivíduos. Essa estimativa está muito abaixo dos 7,2 bilhões que somos atualmente. O que acontece? Essa é uma estimativa do valor de equilíbrio. Os seres humanos modernos são um grupo jovem, de apenas aproximadamente 150.000 anos de idade. Ao longo dos últimos 150.000 anos, nossa população cresceu dramaticamente na medida em que preenchemos o globo, mas a mutação é um processo lento e, assim, a diversidade genética não foi mantida e a população humana não está em equilíbrio. O tamanho da população de 10.498 representa uma estimativa de nosso tamanho histórico, ou quantos membros em idade reprodutiva havia há aproximadamente 150.000 anos.

Equilíbrio entre mutação e seleção

As frequências alélicas também podem alcançar um equilíbrio estável quando a introdução de novos alelos por mutação repetida é balanceada por sua remoção por seleção natural. Esse equilíbrio provavelmente explica a persistência de doenças genéticas como polimorfismos de nível baixo em populações humanas. Novas mutações deletérias estão constantemente surgindo de modo espontâneo. Essas mutações podem ser completamente recessivas, ou parcialmente dominantes. A seleção remove-as da população, mas existe um equilíbrio entre o seu aparecimento e a sua remoção.

Iniciaremos com o caso mais simples — a frequência de um recessivo deletério quando é alcançado um equilíbrio entre mutação e seleção. Para essa finalidade, é conveniente expressar o valor adaptativo relativo em termos do **coeficiente de seleção (*s*)**, que é a desvantagem seletiva de (ou a perda de adaptabilidade em) um genótipo:

$W_{A/A}$	$W_{A/a}$	$w_{a/a}$
1	1	$1 - s$

Em seguida, conforme demonstrado no Quadro 18.8, a equação em relação à

frequência de equilíbrio de um alelo recessivo deletério é:

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

Quadro 18.8 Equilíbrio entre seleção e mutação.

Se deixamos q ser a frequência do alelo deletério a e $p = 1 - q$ ser a frequência do alelo normal A, a alteração na frequência alélica em virtude da taxa de mutação μ então é:

$$\Delta q_{mut} = \mu p$$

Um modo simples de expressar os valores adaptativos dos genótipos no caso de um alelo recessivo deletério a é $w_{A/A} = w_{A/a} = 1{,}0$ e $w_{a/a} = 1 - s$, em que s, o coeficiente de seleção, é a perda de adaptabilidade nos homozigotos recessivos. Agora podemos substituir esses valores adaptativos em nossa expressão geral pela alteração da frequência alélica (ver Quadro 18.7) e obter:

$$\Delta q_{\text{sel}} = \frac{-pq(sq)}{1 - sq^2} = \frac{-spq^2}{1 - sq^2}$$

O equilíbrio significa que o aumento na frequência alélica em virtude de mutação equilibra exatamente a diminuição na frequência alélica em virtude da seleção; assim:

$$\mu\hat{p} = \frac{-s\hat{p}\hat{q}^2}{1 - s\hat{q}^2}$$

A frequência de um alelo recessivo deletério (q) no equilíbrio será consideravelmente pequena, então $1 - sq^2 \approx 1$, e temos:

$$\mu\hat{p} = -s\hat{p}\hat{q}^2$$

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

no equilíbrio.

Essa equação demonstra que a frequência no equilíbrio depende da razão μ/*s*. Quando a taxa de mutação de $A \rightarrow a$ se tornar maior e a desvantagem seletiva se tornar menor, a frequência no equilíbrio ($\hat{q}$) de um alelo recessivo deletério aumentará. Como um exemplo, um alelo letal recessivo ($s = 1$) que surge por meio de mutação do alelo selvagem na taxa de $\mu = 10^{-6}$ apresentará uma frequência de equilíbrio de 10^{-3}.

Consideraremos o equilíbrio entre a seleção e a mutação em relação ao caso discretamente mais complicado de um alelo deletério parcialmente dominante — ou seja, um alelo com algum efeito deletério em heterozigotos, bem como seu efeito em homozigotos. Definiremos *h* como o grau de dominância do alelo deletério. Quando *h* é 1, o alelo deletério é totalmente dominante, e quando *h* é 0, o alelo deletério é totalmente recessivo. Em seguida, os valores adaptativos são

$W_{A/A}$	$W_{A/a}$	$w_{a/a}$
1	$1 - hs$	$1 - s$

em que *a* é um alelo deletério parcialmente dominante. Uma derivação semelhante àquela no Quadro 18.8 nos fornece:

$$\hat{q} = \frac{\mu}{hs}$$

Aqui está um exemplo. Se $\mu = 10^{-6}$ e o alelo letal não for totalmente recessivo, mas causar uma redução de 5% na adaptabilidade dos heterozigotos ($s = 1{,}0$, $h = 0{,}05$), então:

$$\hat{q} = \frac{\mu}{hs} = 2 \times 10^{-5}$$

Esse resultado é menor em duas ordens de magnitude do que a frequência de equilíbrio em relação ao caso puramente recessivo descrito anteriormente. Em

geral, então, podemos esperar que alelos deletérios completamente recessivos apresentem frequências muito mais altas do que aquelas de alelos parcialmente dominantes, tendo em vista que os alelos recessivos estão protegidos nos heterozigotos.

> CONCEITOS-CHAVE A quantidade de variação genética nas populações representa um equilíbrio entre forças opostas: mutação e migração, que adicionam nova variação, *versus* deriva e seleção, que removem a variação. A seleção balanceadora também atua para manter a variação nas populações. Como resultado desses processos, as frequências alélicas podem alcançar valores de equilíbrio, explicando o motivo pelo qual as populações com frequência mantêm altos níveis de variação genética.

18.6 Aplicações biológicas e sociais

Assim como os princípios da física orientam os engenheiros que projetam pontes e jatos, também assim os princípios da genética de populações tocam todas as nossas vidas de muitas maneiras, despercebidas. No Capítulo 19, você verá como a genética de populações figura de modo proeminente na busca por genes que contribuem para o risco de doenças em pessoas, com a utilização de conceitos tais como o desequilíbrio de ligação descrito neste capítulo. Nesta seção final do capítulo, examinaremos quatro outras áreas nas quais os princípios da genética de populações estão sendo aplicados em questões que afetam as sociedades modernas.

Genética da conservação

Biólogos conservacionistas que estão tentando salvar espécies selvagens ameaçadas, e profissionais de zoológicos que estão tentando manter pequenas populações de animais em cativeiro, com frequência realizam análises de genética de populações. Anteriormente, discutimos como um gargalo genético causou perda de variação genética no condor-da-califórnia e um aumento na

frequência de uma forma letal de nanismo. Os gargalos também podem aumentar o nível de endocruzamento em uma população, levando, talvez, a um declínio na adaptabilidade por meio da depressão por endocruzamento. Entretanto, a questão é complexa, tendo em vista que o endocruzamento nem sempre está associado a um declínio na adaptabilidade. O endocruzamento por vezes pode ajudar a remover alelos recessivos deletérios de uma população. A seleção purificadora é mais eficaz na eliminação de alelos recessivos deletérios, tendo em vista que a classe homozigota recessiva se torna mais frequente nas populações endocruzadas. Portanto, biólogos conservacionistas têm debatido se devem tentar maximizar a diversidade genética e minimizar o endocruzamento, ou submeter deliberadamente as populações de zoológicos ao endocruzamento com o objetivo de remover os alelos deletérios.

Para ajudar a abordar essa questão, pesquisadores procuraram por evidências de sucesso na remoção entre populações de zoológicos. Definiremos a depressão por endocruzamento como delta (δ):

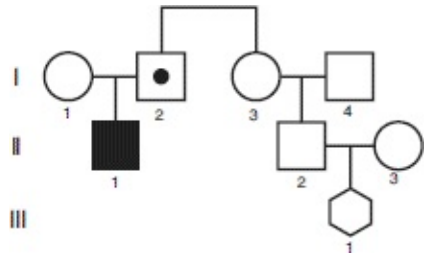
$$\delta = 1 - \frac{w_f}{w_0}$$

em que w_f é a adaptabilidade de indivíduos endocruzados e w_0 é o valor adaptativo de indivíduos não endocruzados. O valor de δ será positivo quando houver um declínio na adaptabilidade com o endocruzamento, mas negativo quando a adaptabilidade melhorar com o endocruzamento. Pesquisadores calcularam δ para 119 populações de zoológico, incluindo 88 espécies, e observaram evidências de que a remoção havia melhorado a adaptabilidade (valores negativos para δ) em 14 populações. Ainda assim, não está claro se o endocruzamento deliberado de animais de zoológico é recomendável. Por um lado, embora 14 das 119 populações tenham melhorado, a maioria das populações declinou na adaptabilidade quando foi endocruzada. Portanto, se iniciarmos com uma pequena população de zoológico e realizarmos propositalmente o endocruzamento dos animais, um declínio na adaptabilidade será o desfecho mais provável.

Cálculo dos riscos de doenças

No Capítulo 2, vimos como os alelos de distúrbios genéticos podiam ser traçados em heredogramas e discutimos como calcular o risco de que um casal tenha um filho que herde tal distúrbio. Os princípios da genética de populações nos possibilitam estender esse tipo de análise. Consideraremos dois exemplos.

O alelo em relação à fibrose cística (FC) ocorre a uma frequência de aproximadamente 0,025 em caucasianos. No heredograma de uma família caucasiana a seguir, o indivíduo II-2 apresenta um primo em primeiro grau (II-1) com fibrose cística. II-2 é casado com uma caucasiana não relacionada (II-3), e eles estão planejando ter um filho. Qual é a chance de que o filho (III-1) apresente fibrose cística?



Um dos avós maternos de II-2 deve ter sido um portador. Começamos calculando a probabilidade de que III-1 herde esse alelo da fibrose cística desse avô ou dessa avó através de seu pai, II-2, utilizando métodos já familiares do Capítulo 2. A probabilidade de que esse avô ou essa avó tenha transmitido o alelo da doença para I-3 é 1/2. A probabilidade de que I-3 o tenha transmitido para II-2 e de que II-2 o transmita para III-1 também são, ambas, 1/2. Assim, a probabilidade de III-1 herdar o alelo da FC de II-2 é $(\frac{1}{2})^3$, ou 1/8. Agora estendemos o cálculo para determinar a probabilidade de que III-1 possa herdar o alelo da fibrose cística de sua mãe, II-3. O indivíduo II-3 não apresenta FC, mas não temos certeza se ela é ou não uma portadora. Se a frequência (q) do alelo da doença na população é de 0,025, então a probabilidade de que um indivíduo não afetado, tal como II-3, seja um portador é de $2pq/(1 - q^2) = 0,049$. Se II-3 for portadora, existe então uma chance de $\frac{1}{2}$ de que ela transmita o alelo da doença para III-1. Essas são todas probabilidades independentes; assim, podemos utilizar a regra do produto. A

probabilidade de que III-1 apresente fibrose cística é:

$$\frac{1}{8} \times \frac{1}{2} \times 0{,}049 = 0{,}003$$

A frequência de fibrose cística entre caucasianos é $p^2 = (0{,}025)^2 = 0{,}000625$. Esses cálculos nos informam que os indivíduos que têm um primo em primeiro grau com fibrose cística apresentam um risco 0,003 ÷ 0,000625 = 4,9 vezes mais alto de ter um filho com a doença do que membros da população em geral.

Aqui está outra aplicação da genética de populações para avaliar o risco de doenças. A anemia falciforme, uma doença recessiva, apresenta uma frequência de aproximadamente 0,25%, ou 1 em 400, entre afro-americanos (ver Capítulo 6). Aplicando a lei de Hardy-Weinberg, estimamos a frequência do alelo da doença (Hb^S) como 0,05. Qual seria a frequência esperada dessa doença entre a descendência de afro-americanos que são primos de primeiro grau? Utilizando o método descrito no Quadro 18.2, calculamos que o coeficiente de endocruzamento (F) em relação à descendência de casamentos entre primos de primeiro grau é de 1/16. Na seção anterior sobre o endocruzamento, vimos que a frequência de homozigotos aumenta quando existe endocruzamento, conforme demonstrado por essa equação:

$$f_{a/a} = q^2 + pqF$$

Com a utilização dessa equação, obtemos:

$$f(Hb^S/Hb^S) = (0{,}05)^2 + (0{,}05 \times 0{,}95)\frac{1}{16} = 0{,}0055$$

Isso representa um aumento de 2,2 vezes no risco de ter um filho com a doença para casamentos de primos de primeiro grau, em comparação àquele em um casamento entre indivíduos não relacionados.

DNA forense

Criminosos podem deixar evidências de DNA na cena de um crime na forma de

sangue, sêmen, cabelos/pelos ou até mesmo células bucais da saliva em uma ponta de cigarro. A reação da cadeia de polimerase (PCR) possibilita que cientistas forenses amplifiquem quantidades muito pequenas de DNA e determinem o genótipo do indivíduo que deixou a amostra. Se o DNA encontrado em uma cena de crime corresponder àquele do suspeito, então eles "podem ser" o mesmo indivíduo. A frase-chave aqui é "podem ser", e é aqui que a genética de populações realiza o seu papel. Vejamos como isso funciona.

Considere dois *loci* de microssatélites, cada um com múltiplos alelos: A_1, A_2… A_n e B_1, B_2… B_n. Cientistas forenses determinam que uma amostra de DNA de uma cena de crime e uma do suspeito são ambas A_3/A_8 B_1/B_7. Eles determinaram que existe uma "correspondência" entre a evidência e o suspeito. A correspondência comprova que a evidência do DNA é originária do suspeito? Ela prova que o suspeito estava na cena do crime?

O que os geneticistas de populações fazem com esse tipo de evidência é testar uma hipótese específica: *a evidência é originária de outra pessoa além do suspeito.* Isso é o que os estatísticos denominam "hipótese nula", ou a hipótese que é considerada verdadeira, exceto se evidências demonstrarem que ela é muito improvável (ver Capítulo 4). Para realizar o teste, calculamos a probabilidade de observar uma correspondência entre a evidência e o suspeito, dado que o suspeito e a pessoa que deixou a evidência são indivíduos diferentes. Simbolicamente, escrevemos:

$$\text{Prob (correspondência} \mid \text{indivíduos diferentes)}$$

em que "|" significa "dado que". Se essa probabilidade for muito pequena, podemos então rejeitar a hipótese nula e argumentar em favor de uma hipótese alternativa: *a evidência foi deixada pelo suspeito.* Nunca comprovamos formalmente que o suspeito deixou a evidência, tendo em vista que pode haver hipóteses alternativas, tais como: *a evidência foi deixada pelo gêmeo idêntico do suspeito*.

Para calcular a probabilidade de observar uma correspondência entre a evidência e o suspeito se a evidência for originária de um indivíduo diferente,

precisamos conhecer as frequências dos alelos de microssatélites na população.

A_4	0,03
A_6	0,05
B_1	0,01
B_7	0,12

A prob (correspondência | indivíduos diferentes) é a mesma probabilidade de que a evidência seja originária de um indivíduo escolhido aleatoriamente. Podemos calcular essa probabilidade utilizando as frequências alélicas anteriores. Primeiramente, presumiremos que a lei de Hardy-Weinberg seja aplicável e calcularemos a probabilidade de ser A_4/A_6 no primeiro *locus* e B_1/B_7 no segundo:

Prob $(A_4/A_6) = 2pq = 2 \times 0{,}03 \times 0{,}05 = 0{,}003$
Prob $(B_1/B_7) = 2 \times 0{,}01 \times 0{,}12 = 0{,}0024$

Para combinar essas duas probabilidades, precisamos realizar mais uma presunção. Precisamos presumir que os dois *loci* são *independentes*; ou seja, que os *loci* estão em *equilíbrio de ligação*. Ao realizar essa presunção, podemos aplicar a *regra do produto* a eventos independentes (ver Capítulo 2) e determinar que:

$$\text{Prob (correspondência | indivíduos diferentes)} = \text{Prob } (A_4/A_6) \times \text{Prob } (B_1/B_7) = 7{,}2 \times 10^{-6}$$

Portanto, a probabilidade sob a hipótese nula de que a evidência seja originária de alguma outra pessoa além do suspeito é de $7{,}2 \times 10^{-6}$, ou aproximadamente 7 em um milhão. Essa é uma probabilidade pequena e, assim, a hipótese nula aparenta ser improvável nesse caso. Entretanto, se a Prob (correspondência | indivíduos diferentes) fosse 0,1, 10% da população seria correspondente e poderia ter deixado a evidência. Naquele caso, não rejeitaríamos a hipótese nula.

Dois microssatélites não proporcionam muito poder de discriminação, e então o FBI, nos EUA, utiliza um conjunto de 13 microssatélites. Os *loci* de microssatélites apresentam tipicamente grandes números de alelos (10 a 20, ou mais); portanto, o número de possíveis genótipos com base em 13 microssatélites é astronomicamente grande. Com 10 alelos por *locus*, existem 55 possíveis genótipos em cada *locus* e 55^{13}, ou $4{,}2 \times 10^{22}$, possíveis genótipos *multilocus* para 13 *loci*. O FBI também reuniu uma base de dados denominada CODIS (Combined DNA Index System) que contém as frequências de diferentes alelos nesses *loci* na população, incluindo dados específicos de diferentes grupos étnicos e regiões do país.

Procura no Google por seus parceiros de DNA

Neste capítulo, revisamos os princípios básicos da genética de populações e discutimos muitas aplicações para a genética humana. A teoria básica da genética de populações tem estado presente há quase 100 anos, mas apenas na última década o desenvolvimento de tecnologias com base no alto processamento do DNA para a genotipagem de indivíduos trouxe à tona os complexos padrões de variação entre e dentro das populações humanas. Não apenas tem sido possível desembaraçar muitos dos detalhes a respeito de como e quando os seres humanos povoaram o globo a partir da sua origem na África, como também os geneticistas obtiveram uma profunda compreensão sobre como forças tais como a seleção natural e a deriva genética modelaram quem somos.

O que o futuro reserva para nós? Logo o sequenciamento de um genoma humano poderá custar um pouco mais do que uma nova bicicleta. Uma estudante universitária poderá esfregar o interior de sua bochecha com um *Q-tip* e depositar a amostra em um quiosque enquanto ouve música. Semanas depois, ela poderá visualizar a sequência de seu genoma em um *site*, compará-la àquela de seus amigos e parentes, e aprender a respeito de seus ancestrais. Conforme veremos no próximo capítulo, a nossa capacidade de prever os riscos de doenças, os talentos e outros traços de uma pessoa a partir do seu genótipo está melhorando. Até a medida na qual o gosto de uma pessoa pela música ou a admiração por esportes radicais apresenta fundamentos genéticos, teoricamente uma pessoa poderia

“buscar no Google” parceiros de DNA que provavelmente compartilham os mesmos interesses. A tecnologia do DNA e a teoria da genética de populações já estão funcionando; entretanto, existem questões sociais e éticas a serem abordadas. As informações poderão ser mantidas em sigilo, e como? Existem muitos limites sobre o que uma pessoa deve saber a respeito de sua própria sequência? O governo deve sequenciar o genoma de qualquer pessoa quando ela nasce? Planos de saúde podem exigir que seus clientes submetam as sequências do seu genoma? Uma compreensão sobre a ciência pode auxiliar na determinação de como essas questões são respondidas.

RESUMO

A genética de populações procura compreender as leis que regulam e as forças que influenciam a quantidade de variação genética nas populações e as alterações na variação genética ao longo do tempo. O conceito do *pool* gênico fornece um modelo para pensarmos a respeito da transmissão da variação genética de uma geração para a próxima em relação a uma população inteira. A teoria básica da genética de populações tem início com uma população idealizada de tamanho infinito e na qual o cruzamento é aleatório. Em uma referida população, a lei de Hardy-Weinberg define a relação entre as frequências alélicas no *pool* gênico e as frequências genotípicas na população.

As populações reais normalmente se desviam, em pequeno ou grande grau, do modelo de Hardy-Weinberg. Uma fonte de desvio advém do tipo de cruzamento não aleatório ou preferencial. Se os indivíduos cruzam preferencialmente com outros que compartilham um fenótipo semelhante, então haverá um excesso de homozigotos para genes que controlam aquele fenótipo em comparação às expectativas de Hardy-Weinberg. Quando os indivíduos cruzam mais frequentemente com parentes do que o esperado ao acaso, haverá então um excesso de genótipos homozigotos em todo o genoma e a população se torna endocruzada. Até mesmo quando as populações locais de uma espécie estão de acordo com as expectativas de Hardy-Weinberg, aquelas estão aptas a ser isoladas de outras populações em locais distantes. Portanto, uma espécie com

frequência é composta por uma série de subpopulações geneticamente distintas; ou seja, as espécies demonstram estrutura genética de população.

Diversas forças podem adicionar novas variações a uma população ou remover dela a variação existente. A mutação é a fonte definitiva de toda a variação genética. Geneticistas de populações determinaram estimativas razoavelmente precisas da taxa na qual surgem novas mutações nas populações. A migração também pode trazer nova variação para uma população. A migração resulta em alguns indivíduos que são geneticamente misturados, que apresentam ancestrais de múltiplas populações. A recombinação genética também pode adicionar variação às populações por meio da recombinação de alelos em novos haplótipos.

Duas forças controlam o destino da variação genética nas populações. Primeiramente, a deriva genética é uma força aleatória que pode levar à perda ou à fixação de um alelo como resultado do erro de amostragem em populações finitas. A deriva age com mais força sobre populações pequenas e menor sobre populações grandes. Em segundo lugar, a seleção natural direciona alterações nas frequências alélicas nas populações ao longo do tempo. Os alelos que intensificam a adaptabilidade dos indivíduos que os carregam aumentarão em frequência e poderão se tornar fixos, enquanto alelos deletérios que reduzem a adaptabilidade serão removidos da população.

O objetivo fundamental da genética de populações é compreender as contribuições relativas dos sistemas de cruzamento, mutação, migração, recombinação, deriva e seleção natural para a quantidade e a distribuição da variação genética nas populações. Neste capítulo, vimos como pesquisas em genética de populações desenvolveram a teoria básica e coletaram uma vasta quantidade de dados para a conquista desse objetivo. Nossa compreensão sobre a genética de populações de nossa própria espécie é notavelmente detalhada.

Finalmente, os métodos e os resultados da genética de populações nos informam a respeito do processo evolutivo e apresentam aplicações práticas para questões que as sociedades modernas enfrentam. A teoria e as análises da genética de populações desempenham papéis importantes no manejo de espécies ameaçadas, na identificação de criminosos, no cultivo de plantas e na criação

animal, e na avaliação dos riscos de um casal ter um filho com uma doença.

TERMOS-CHAVE

adaptabilidade absoluta
adaptabilidade darwiniana
adaptabilidade relativa
adaptação
alelo neutro
coeficiente de endocruzamento
coeficiente de seleção (s)
cruzamento preferencial negativo
cruzamento preferencial positivo
depressão por endocruzamento
deriva genética aleatória
desequilíbrio de ligação (LD)
diversidade gênica (GD)
diversidade nucleotídica
efeito fundador
endocruzamento
equilíbrio de Hardy-Weinberg
equilíbrio de ligação
estrutura populacional
evolução neutra
fixado
fluxo gênico
frequência alélica
frequência genotípica
gargalo
genética de populações
haplótipo
HapMap

heterozigosidade (H)
idêntico por descendência (IBD)
isolamento pela distância
lei de Hardy-Weinberg
locus
microssatélite
migração
mistura genética
número de haplótipos (NH)
painel de descoberta
polimorfismo de nucleotídio único (SNP)
pool gênico
população
rede de haplótipos
relógio molecular
seleção artificial
seleção balanceadora
seleção direcional
seleção natural
seleção positiva
seleção purificadora
sítios segregantes (S)
SNP comum
SNP raro
taxa de mutação (μ)

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Aproximadamente 70% de todos os caucasianos conseguem degustar a substância química feniltiocarbamida, e o restante não consegue. A capacidade de degustar essa substância química é determinada pelo alelo dominante *T*, e a incapacidade de degustar é determinada pelo alelo

recessivo *t*. Se presumirmos que a população está em equilíbrio de Hardy-Weinberg, quais são as frequências genotípicas e alélicas nessa população?

Solução

Tendo em vista que 70% são degustadores (*T/T* e *T/t*), 30% devem ser não degustadores (*t/t*). Essa frequência de homozigotos recessivos é igual a q^2; assim, para obter *q*, simplesmente calculamos a raiz quadrada de 0,30:

$$q = \sqrt{0{,}30} = 0{,}55$$

Tendo em vista que $p + q = 1$, podemos escrever $p = 1 - q = 1 - 0{,}55 = 0{,}45$.

Agora podemos calcular:

$$p^2 = (0{,}45)^2 = 0{,}20, \quad \text{a frequência de } T/T$$

$$2pq = 2 \times 0{,}45 \times 0{,}55 = 0{,}50, \quad \text{a frequência de } T/t$$

$$q^2 = 0{,}3, \quad \text{a frequência de } t/t$$

Problema resolvido 2. Em uma grande população experimental de *Drosophila*, o valor adaptativo relativo de um fenótipo recessivo é calculado como 0,90, e a taxa de mutação em relação ao alelo recessivo é 5×10^{-5}. Se a população estiver em equilíbrio, quais frequências alélicas podem ser previstas?

Solução

Aqui, a mutação e a seleção estão atuando em direções opostas e, assim, é previsto um equilíbrio. Tal equilíbrio é descrito pela fórmula:

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

Na presente questão:

$$\mu = 5 \times 10^{-5} \text{ e } s = 1 - w = 1 - 0{,}9 = 0{,}1$$

Portanto:

$$\hat{q} = \sqrt{\frac{5 \times 10^{-5}}{0{,}1}} = 0{,}022$$

$$\hat{p} = 1 - 0{,}022 = 0{,}978$$

Problema resolvido 3. Uma colônia de 50 papagaios-do-mar (*Fratercula corniculata*) é estabelecida em um zoológico e ali é mantida durante 30 gerações.

a. Se o coeficiente de endocruzamento dos membros fundadores for zero (F = 0,0), qual é o coeficiente de endocruzamento esperado para essa população atualmente?

b. Em relação a um alelo deletério para doença, com uma frequência de 0,001 na natureza, qual é a frequência prevista de aves homozigotas afetadas na natureza e na população do zoológico atualmente?

Solução

a. No Quadro 18.3, vimos que o endocruzamento aumentará como uma função do tamanho da população (N) ao longo do tempo (t), conforme medido nas gerações, de acordo com a equação a seguir:

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t (1 - F_0)$$

Substituindo por N = 50, t = 30 e F_0 = 0, obtemos:

$$F_{30} = 1 - \left(1 - \frac{1}{2 \times 50}\right)^{30} (1 - 0) = 0{,}26$$

b. Se a frequência de um alelo de doença recessiva (q) na natureza é 0,001, então por meio da aplicação da lei de Hardy-Weinberg prevemos que a frequência de

indivíduos homozigotos afetados na natureza será $q^2 = 10^{-6}$. Em relação à população do zoológico, a frequência de homozigotos será mais alta em virtude do endocruzamento, de acordo com a equação a seguir:

$$f_{a/a} = q^2 + pqF$$

Substituindo por $q = 0{,}001$, $p = 0{,}999$ e $F = 0{,}26$, obtemos:

$$f_{a/a} = 10^{-6} + (0{,}001 \times 0{,}999 \times 0{,}26) = 2{,}61 \times 10^{-4}$$

A razão de $2{,}61 \times 10^{-4}$ a 10^{-6} nos demonstra que existe um aumento de 261 vezes na frequência esperada de indivíduos afetados na população do zoológico atual em comparação à população selvagem ancestral.

Problema resolvido 4. Em um julgamento penal, o promotor apresenta os genótipos em relação a três *loci* de microssatélites do conjunto CODIS do FBI. Ele relata que uma amostra de DNA da cena do crime e uma do suspeito apresentam, ambas, o genótipo FGA_1/FGA_4, $TPOX_1/TPOX_3$, VWA_2/VWA_7 nesses três microssatélites. Ele também apresenta as frequências alélicas em relação à população em geral à qual o suspeito pertence (ver tabela a seguir). Qual é a probabilidade de que o genótipo da evidência de DNA corresponda àquele do suspeito, dado que a pessoa que cometeu o crime e o suspeito sejam indivíduos diferentes? Quais presunções você realiza quando calcula essa probabilidade?

Alelo	Frequência
FGA_1	0,30
FGA_4	0,26
$TPOX_1$	0,32
$TPOX_3$	0,65
VWA_2	0,23
VWA_7	0,59

Solução

A probabilidade de que o genótipo da evidência de DNA corresponda àquele do suspeito, dado que a pessoa que cometeu o crime e o suspeito sejam indivíduos diferentes, é a mesma probabilidade de que um membro da população escolhido aleatoriamente apresente o mesmo genótipo da evidência de DNA. A probabilidade de uma pessoa escolhida aleatoriamente ser $FGA_1/FGA_4 = 2pq = 2$ (0,30) (0,26) = 0,156 e, de modo semelhante, a probabilidade de uma pessoa aleatória ser $TPOX_1/TPOX_3$ = 0,416 e VWA_2/VWA_7 = 0,2714. Aplicando a regra da multiplicação, a probabilidade de um membro aleatório da população ser FGA_1/FGA_4, $TPOX_1/TPOX_3$, VWA_2/VWA_7 = 0,156 × 0,416 × 0,2714 = 0,0176. Ao calcular essa probabilidade, presumimos que a população está em equilíbrio de Hardy-Weinberg e que os três *loci* em questão estão em equilíbrio de ligação entre si.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Qual indivíduo na Figura 18.3 apresenta o maior número de *loci* heterozigotos, e qual indivíduo apresenta o menor número?

2. Suponha que os setes cromossomos na Figura 18.4 A representem uma amostra aleatória de cromossomos de uma população.

a. Calcule a diversidade genética (*GD*) em separado em relação ao indel, o *locus* de microssatélite e o SNP na posição 3.

b. Se a sequência foi encurtada de modo que você tivesse os dados apenas em relação às posições 1 a 24, quantos haplótipos haveria ali?

c. Calcule o parâmetro do desequilíbrio de ligação (*D*) entre os SNP nas posições 29 e 33.

3. Observando a Figura 18.6, você consegue contar quantos haplótipos mitocondriais foram transportados da Ásia para as Américas?

4. Na Figura 18.13, a coluna “não relacionados” (*azul*) em relação ao Japão é mais alta do que a coluna “não relacionados” em relação à França. O que isso lhe informa?

5. Na Figura 18.14, alguns indivíduos apresentam alelos de SNP únicos — por exemplo, o alelo T no SNP4 ocorre apenas no indivíduo 12. Você consegue identificar dois indivíduos que apresentem, cada um, alelos únicos em dois SNP?

6. Observando a Figura 18.20, as pessoas do Oriente Médio tendem a apresentar níveis mais altos ou mais baixos de heterozigosidade em comparação às pessoas da Ásia Oriental? Por que esse pode ser o caso?

PROBLEMAS BÁSICOS

7. Quais são as forças que podem alterar a frequência de um alelo em uma população?

8. Quais presunções são feitas quando se utiliza a fórmula de Hardy-Weinberg para estimar as frequências genotípicas a partir das frequências alélicas?

9. Em uma população de camundongos, existem dois alelos no *locus A* (A_1 e A_2). Testes demonstraram que, nessa população, existem 384 camundongos de genótipo A_1/A_1, 210 A_1/A_2 e 260 de A_2/A_2. Quais são as frequências dos dois alelos na população?

10. Em uma população natural de *Drosophila melanogaster*, o gene da desidrogenase alcoólica apresenta dois alelos, denominados *F* (rápido) e *S* (lento), com as frequências de *Adh-F* em 0,75 e *Adh-S* em 0,25. Em uma amostra de 480 moscas dessa população, quantos indivíduos de cada classe genotípica você esperaria observar sob o equilíbrio de Hardy-Weinberg?

11. Em uma população de laboratório de *Drosophila* de cruzamento aleatório, 4% das moscas apresentam corpo preto (codificado pelo autossômico recessivo *b*) e 96% apresentam corpo marrom (o tipo selvagem, codificado por *B*). Se presumirmos que essa população está em equilíbrio de Hardy-Weinberg, quais são as frequências alélicas de *B* e *b* e as frequências genotípicas de *B/B* e *B/b*?

12. Em uma população de uma espécie de besouro, você observa que existe uma razão de 3:1 de coberturas brilhantes e foscas das asas. Essa razão

comprova que o alelo *brilhante* é dominante? (Presuma que os dois estados são causados por dois alelos de um gene.) Caso negativo, o que ela comprova? Como você elucidaria a situação?

13. A fibrose cística (FC) é um distúrbio autossômico recessivo que ocorre relativamente com frequência entre pessoas de descendência europeia. Em uma comunidade Amish em Ohio, pesquisadores médicos relataram a ocorrência de fibrose cística (FC) como sendo de 1/569 nascimentos vivos. Com a utilização da regra de Hardy-Weinberg, estime a frequência de portadores do alelo da doença nessa população Amish.

14. Os valores adaptativos relativos de três genótipos são $w_{A/A} = 1{,}0$, $w_{A/a} = 1{,}0$, e $w_{a/a} = 0{,}7$.

a. Se a população tem início com a frequência alélica $p = 0{,}5$, qual é o valor de p na próxima geração?

b. Qual é a frequência alélica no equilíbrio prevista se a taxa de mutação de *A* para *a* é 2×10^{-5}?

15. Indivíduos *A/A* e *A/a* são igualmente férteis. Se 0,1% da população é *a/a*, qual pressão de seleção existe contra *a/a* se a taxa de mutação de $A \rightarrow a$ é 10^{-5}? Presuma que as frequências alélicas estão em seus valores de equilíbrio.

16. Quando os alelos em um *locus* atuam de modo semidominante sobre a adaptabilidade, o valor adaptativo relativo do heterozigoto está no meio da distância entre as duas classes homozigotas. Por exemplo, genótipos com semidominância no *locus A* podem apresentar esses valores adaptativos relativos: $w_{A/A} = 1{,}0$, $w_{A/a} = 0{,}9$ e $w_{a/a} = 0{,}8$.

a. Altere um desses valores adaptativos de modo que *a/a* se torne um alelo recessivo deletério.

b. Altere um desses valores adaptativos de modo que *A/A* se torne um alelo dominante favorecido.

17. Se o alelo recessivo em relação a uma doença recessiva ligada ao X em seres humanos apresenta uma frequência de 0,02 na população, qual proporção de indivíduos na população apresentará a doença? Presuma que

a população é 50:50 masculina:feminina.

18. O daltonismo é um distúrbio recessivo ligado ao X em seres humanos, causado por mutações em um dos genes que codificam a proteína opsina sensível à luz. Se o alelo mutante apresenta uma frequência de 0,08 na população, qual proporção de mulheres será portadora? Presuma que a população é 50:50 masculina:feminina.

19. Uma nova mutação neutra apresenta maior probabilidade de alcançar a fixação em uma população grande ou pequena?

20. Aparentemente está claro que o endocruzamento causa uma redução na adaptabilidade. Você pode explicar o motivo?

21. Em uma população de 50.000 indivíduos diploides, qual é a probabilidade de que uma nova mutação neutra finalmente alcance a fixação? Qual é a probabilidade de que ela finalmente seja perdida da população?

22. O endocruzamento em uma população causa um desvio das expectativas de Hardy-Weinberg, de tal modo que existem mais homozigotos do que o esperado. Em relação a um *locus* com um alelo deletério raro a uma frequência de 0,04, qual seria a frequência de homozigotos em relação ao alelo deletério em populações com coeficientes de endocruzamento de F = 0,0 e F = 0,125?

23. A anemia falciforme é um distúrbio autossômico recessivo causado pela substituição de um aminoácido na proteína β-hemoglobina. A mutação no DNA subjacente a essa substituição é um SNP que altera um códon GAG para o aminoácido glutamato para um GTG, que codifica uma valina. A frequência de anemia falciforme entre afro-americanos é de aproximadamente 1/400. Qual é a frequência desse códon GTG no gene da β-hemoglobina entre afro-americanos?

24. Você tem uma amostra de 10 sequências de DNA de 100 pb de comprimento de uma seção de um gene altamente conservado de 10 indivíduos de uma espécie. As 10 sequências são quase totalmente

idênticas; entretanto, cada sequência carrega um SNP único não observado em qualquer uma das outras. Qual é a diversidade nucleotídica em relação a essa amostra de sequências?

PROBLEMAS DESAFIADORES

25. A Figura 18.14 apresenta os dados de haplótipos em relação ao gene *G6PD* em uma amostra mundial de pessoas.

a. Desenhe uma rede de haplótipos em relação a esses haplótipos. Rotule os ramos nos quais cada SNP ocorre.

b. Qual dos haplótipos apresenta a maioria das conexões com outros haplótipos?

c. Em quais continentes esse haplótipo é encontrado?

d. Contando o número de SNP ao longo dos ramos da sua rede, quantas diferenças existem entre os haplótipos 1 e 12?

26. A Figura 18.12 demonstra um heredograma de um cruzamento entre meios-irmãos.

a. Se o coeficiente de endocruzamento em relação ao ancestral comum (A) na Figura 18.12 é 1/2, qual é o coeficiente de endocruzamento de I?

b. Se o coeficiente de endocruzamento do indivíduo I na Figura 18.12 é 1/8, qual é o coeficiente de endocruzamento do ancestral comum, A?

27. Considere 10 populações que apresentam as frequências genotípicas demonstradas na tabela a seguir:

População	*A/A*	*A/a*	*a/a*
1	1,0	0,0	0,0
2	0,0	1,0	0,0

3	0,0	0,0	1,0
4	0,50	0,25	0,25
5	0,25	0,25	0,50
6	0,25	0,50	0,25
7	0,33	0,33	0,33
8	0,04	0,32	0,64
9	0,64	0,32	0,04
10	0,986049	0,013902	0,000049

a. Quais das populações estão em equilíbrio de Hardy-Weinberg?

b. Quais são p e q em cada população?

c. Na população 10, descobre-se que a taxa de mutação de $A \rightarrow a$ é de 5×10^{-6}. Qual deve ser a adaptabilidade do fenótipo *a/a* se a população está em equilíbrio?

d. Na população 6, o alelo *a* é deletério; além disso, o alelo *A* é incompletamente dominante; assim, *A/A* é perfeitamente adaptado, *A/a* apresenta adaptabilidade de 0,8, e *a/a* apresenta adaptabilidade de 0,6. Se não existir mutação, quais serão p e q na próxima geração?

28. O gene da hemoglobina B (*Hb*) apresenta um alelo comum (*A*) de um SNP (*rs334*) que codifica a forma Hb^A da hemoglobina (adulta) e um alelo raro

(*T*) que codifica a forma falciforme da hemoglobina, Hb^S. Entre 571 residentes de uma vila na Nigéria, observou-se que 440 indivíduos eram *A/A* e 129 eram *A/T*, e 2 eram *T/T*. Utilize o teste do χ^2 para determinar se essas frequências genotípicas observadas correspondem às expectativas de Hardy-Weinberg.

29. Uma população apresenta as frequências gaméticas a seguir em dois *loci*: *AB* = 0,4, *Ab* = 0,1, *aB* = 0,1 e *ab* = 0,4. Se possibilitarmos que a população cruze aleatoriamente até que o equilíbrio de ligação seja alcançado, qual será a frequência esperada de indivíduos que são heterozigotos em ambos os *loci*?

30. Duas espécies de palmeiras diferem em 50 pb em um trecho de 5.000 pb de DNA que se acredita ser neutro. A taxa de mutação em relação a essas espécies é de 2×10^{-8} substituições por sítio por geração. O tempo de geração em relação a essas espécies é de 5 anos. Estime o tempo desde que essas espécies apresentavam um ancestral comum.

31. O daltonismo em seres humanos é causado por um alelo recessivo ligado ao X. Dez por cento dos homens de uma população grande e de cruzamento aleatório são daltônicos. Um grupo representativo de 1.000 pessoas dessa população migra para uma ilha do Pacífico Sul, onde já existem 1.000 habitantes e na qual 30% dos homens são daltônicos. Presumindo que o equilíbrio de Hardy-Weinberg é totalmente aplicável (nas duas populações originais antes da migração e na população mista imediatamente após a migração), qual fração de homens e mulheres espera-se que seja daltônica na geração imediatamente após a chegada dos migrantes?

32. Com a utilização de diagramas de heredogramas, calcule o coeficiente de endocruzamento (*F*) em relação à descendência de (a) cruzamentos entre pais e filhos; (b) cruzamentos entre primos em primeiro grau; (c) cruzamentos entre tia e sobrinho ou tio e sobrinha; (d) autofertilização de um hermafrodita.

33. Um grupo de 50 homens e 50 mulheres estabelece uma colônia em uma ilha remota. Após 50 gerações de cruzamentos aleatórios, quão frequente seria

um traço recessivo, se ele apresentasse uma frequência de 1/500 no continente? A população permanece do mesmo tamanho durante as 50 gerações e o traço não apresenta efeito sobre a adaptabilidade.

34. A Figura 18.22 demonstra 10 haplótipos de uma população antes de uma varredura seletiva e outros 10 haplótipos muitas gerações depois, após a ocorrência de uma varredura seletiva em relação a essa região cromossômica. Existem 11 *loci* que definem cada haplótipo, incluindo um com um alelo vermelho, que era o alvo da seleção. Na figura, dois *loci* estão designados *A* e *B*. Cada um desses *loci* apresenta dois alelos: um preto e o outro, cinza. Calcule o parâmetro do desequilíbrio de ligação (*D*) entre *A* e *B*, tanto antes quanto após a varredura seletiva. Qual efeito a varredura seletiva apresentou sobre o nível do desequilíbrio de ligação?

35. A taxa de recombinação (*r*) entre os *loci* ligados *A* e *B* é de 0,10. Em uma população, observamos as frequências haplotípicas a seguir:

AB	0,40
aB	0,10
Ab	0,10
ab	0,40

a. Qual é o nível do desequilíbrio de ligação, conforme medido por *D* na geração atual?
b. Qual será *D* na próxima geração?
c. Qual é a frequência esperada do haplótipo *Ab* na próxima geração?
d. Com a utilização de um *software* de planilhas, construa um gráfico do declínio em *D* ao longo de 10 gerações.

36. O alelo *B* é um autossômico dominante deletério. A frequência de indivíduos afetados é $4{,}0 \times 10^{-6}$. A capacidade reprodutiva desses indivíduos é de aproximadamente 30% daquela dos indivíduos normais.

Estime μ, a taxa na qual *b* sofre mutação para o seu alelo deletério *B*. Presuma que as frequências alélicas estão em seus valores de equilíbrio.

37. Qual é o equilíbrio de heterozigosidade para um SNP em uma população de 50.000 quando a taxa de mutação é 3×10^{-8}?

38. De 31 crianças nascidas de cruzamentos entre pai e filha, 6 morreram no primeiro ano de vida, 12 eram muito anormais e morreram na infância, e 13 eram normais. A partir dessa informação, calcule aproximadamente quantos genes letais recessivos nós apresentamos, em média, em nossos genomas humanos. (Dica: se a resposta for 1, uma filha então apresenta uma chance de 50% de carregar o alelo letal, e a probabilidade de a união produzir uma combinação letal será $1/2 \times 1/4 = 1/8$. Assim, 1 não é a resposta.) Considere também a possibilidade de fatalidades *in utero* não detectadas nos referidos cruzamentos. Como elas afetariam o seu resultado?

39. O *locus B* apresenta dois alelos *B* e *b* com frequências de 0,95 e 0,05, respectivamente, em uma população na geração atual. As adaptabilidades genotípicas nesse *locus* são $w_{B/B} = 1{,}0$, $w_{B/b} = 1{,}0$, e $w_{b/b} = 0{,}0$.

a. Qual será a frequência do alelo *b* em duas gerações?

b. Qual será a frequência do alelo *b* em duas gerações se as adaptabilidades forem $w_{B/B} = 1{,}0$, $w_{B/b} = 0{,}0$ e $w_{b/b} = 0{,}0$?

c. Explique por que existe uma diferença na taxa de alteração em relação à frequência do alelo *b* sob as partes **a** e **b** deste problema.

40. O gene *sd* causa uma doença letal da infância em seres humanos quando em homozigose. Um em 100.000 recém-nascidos morre a cada ano em virtude dessa doença. A taxa de mutação de *Sd* para *sd* é de 2×10^{-4}. Qual deve ser a adaptabilidade do heterozigoto para explicar a frequência do gene observada em vista da taxa de mutação? Atribua um valor adaptativo relativo de 1,0 para os homozigotos *Sd/Sd*. Presuma que a população está em equilíbrio em relação à frequência de *sd*.

41. Se definirmos o *custo da seleção total* para uma população de genes recessivos deletérios como a perda de adaptabilidade por indivíduo afetado (s) multiplicada pela frequência de indivíduos afetados (q^2), então

o custo da seleção = sq^2.

a. Suponha que uma população esteja em equilíbrio entre mutação e deleção em relação a um alelo recessivo deletério, em que $s = 0{,}5$ e $\mu = 10^{-5}$. Qual é a frequência de equilíbrio do alelo? Qual é o custo da seleção?

b. Suponha que iniciamos irradiando membros individuais da população de modo que a taxa de mutação seja duplicada. Qual é a nova frequência de equilíbrio do alelo? Qual é o novo custo de seleção?

c. Se não alterarmos a taxa de mutação, mas em vez disso reduzirmos o coeficiente de seleção para 0,3, o que acontece com a frequência de equilíbrio e o custo da seleção?

42. A seleção balanceadora atua para manter a diversidade genética em um *locus*, tendo em vista que a classe heterozigota apresenta adaptabilidade maior do que as classes homozigotas. Sob esse tipo de seleção, as frequências alélicas na população se aproximam de um ponto de equilíbrio em algum ponto entre 0 e 1. Considere um *locus* com dois alelos *A* e *a* com frequências *p* e *q*, respectivamente. As adaptabilidades genotípicas relativas estão demonstradas a seguir, em que *s* e *g* são as desvantagens seletivas das duas classes homozigotas.

Genótipo	*A/A*	*A/a*	*a/a*
Adaptabilidade relativa	$1 - s$	1	$1 - g$

a. Em equilíbrio, a adaptabilidade média dos alelos *A* (w_A) será igual à adaptabilidade média dos alelos *a* (w_a) (ver Quadro 18.7). Estabeleça a adaptabilidade média dos alelos *A* (w_A) igual à adaptabilidade média dos alelos *a* (w_a). Solucione a equação resultante em relação à frequência do alelo *A*. Essa é a expressão em relação à frequência de equilíbrio de *A* ().

b. Utilizando a expressão que você acabou de derivar, encontre *p* quando $s = 0{,}2$ e $g = 0{,}8$.

CAPÍTULO 19

Herança de Traços Complexos

O ex-astro de basquete Kareem Abdul-Jabbar (2,18 m de altura) e o famoso ex-jóquei Willie Shoemaker (1,50 m) demonstram alguns dos extremos na altura humana – um traço quantitativo. (Associated Press.)

TÓPICOS

19.1 Medida da variação quantitativa

19.2 Modelo genético simples para os traços quantitativos
19.3 Herdabilidade no sentido amplo | Natureza versus criação
19.4 Herdabilidade no sentido restrito | Previsão dos fenótipos
19.5 Mapeamento de QTL em populações com heredogramas conhecidos
19.6 Mapeamento de associação em populações de cruzamento aleatório

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Em relação a qualquer característica em particular, analisar dados para determinar o quanto da variação em uma população ocorre em virtude de fatores genéticos e o quanto em virtude de fatores ambientais
- Utilizar o conhecimento sobre os fenótipos parentais para prever o fenótipo da descendência
- Determinar quantos genes contribuem para a variação genética de um traço
- Identificar os genes específicos que contribuem para a variação em traços quantitativos nas populações.

Observe quase qualquer grande grupo de homens e mulheres e você notará uma variação considerável nas suas alturas – alguns são baixos, alguns altos e alguns têm estatura mediana. Kareem Abdul-Jabbar, um astro pivô de basquete das décadas de 1970 e 1980, era altíssimo, com 2,18 m de altura, enquanto Willie Shoemaker, um famoso jóquei que venceu o Kentucky Derby quatro vezes, tinha meros 1,50 m. Você também pode ter observado que, em algumas famílias, os genitores e seus filhos adultos são todos do grupo dos altos, enquanto em outras famílias, os genitores e os filhos adultos são todos razoavelmente baixos. Tais observações sugerem que os genes desempenham um papel na determinação das nossas alturas. Ainda assim, as pessoas não são segregadas claramente nas categorias alta e baixa, conforme vimos em relação às ervilhas de Mendel. Na primeira inspeção, os traços contínuos, tais como a altura, não aparentam seguir as leis de Mendel, apesar do fato de serem hereditários.

Traços tais como a altura, que demonstram uma amplitude de variação contínua

e que não se comportam do modo mendeliano simples, são conhecidos como **traços quantitativos** ou **complexos**. O termo *traço complexo* com frequência é preferido em virtude da variação em relação a tais traços ser regulada por um "complexo" de fatores genéticos e ambientais. Quão alto você é pode ser parcialmente explicado pelos genes que herdou de seus pais e parcialmente por fatores ambientais, tais como quão bem-nutrido você foi quando criança. Discriminar as contribuições genéticas e ambientais em relação a um fenótipo individual é um desafio substancial, mas os geneticistas apresentam um poderoso conjunto de recursos para a sua conquista.

No início do século 20, quando as leis de Mendel foram redescobertas, surgiu uma controvérsia sobre o fato de elas serem ou não aplicáveis aos traços contínuos. Um grupo conhecido como os biometristas descobriu que existem correlações entre os parentes em relação aos traços contínuos, de modo que genitores altos tendem a ter filhos altos. Entretanto, os biometristas não observaram evidências de que os referidos traços seguissem as leis de Mendel. Alguns biometristas concluíram que os *loci* de Mendel não controlam os traços contínuos. Por outro lado, alguns adeptos do mendelismo pensavam que a variação contínua não era importante e que ela poderia ser ignorada no estudo da herança. Em 1920, essa controvérsia foi resolvida com a formulação da **hipótese multifatorial**. Essa hipótese propôs que os traços contínuos são regulados por uma combinação de múltiplos *loci* mendelianos, cada um com um pequeno efeito sobre o traço e por fatores ambientais. A hipótese multifatorial trouxe os traços quantitativos para o reino da genética mendeliana.

Embora a hipótese multifatorial tenha fornecido uma explicação sensível em relação à variação contínua, a análise mendeliana clássica é inadequada para o estudo dos traços complexos. Se a progênie não puder ser classificada em categorias com proporções esperadas, a abordagem mendeliana então apresenta pouca utilidade para a análise de traços complexos. Em resposta a esse problema, os geneticistas desenvolveram um conjunto de modelos matemáticos e métodos estatísticos para a análise de traços complexos. Por meio da aplicação desses métodos analíticos, os geneticistas realizaram grandes avanços na compreensão sobre os traços complexos. A subárea da genética que desenvolve e aplica esses

métodos para a compreensão da herança de traços complexos é denominada **genética quantitativa**.

No cerne da área da genética quantitativa está o objetivo de definir a **arquitetura genética** de traços complexos. A arquitetura genética é uma descrição de todos os fatores genéticos que influenciam um traço. Ela inclui o número de genes que afetam o traço e a contribuição relativa de cada gene. Alguns genes podem apresentar um grande efeito sobre o traço, enquanto outros apresentam apenas um pequeno efeito. Conforme veremos neste capítulo, a arquitetura genética é a propriedade de uma população específica e pode variar entre as populações de uma espécie. Por exemplo, a arquitetura genética de um traço tal como a pressão arterial sistólica em seres humanos varia entre as diferentes populações. Isso ocorre porque diferentes alelos segregam em diferentes populações e diferentes populações vivenciam ambientes diferentes; portanto, diferentes populações estão aptas a apresentar diferentes arquiteturas em relação a muitos traços.

A compreensão da herança de traços complexos é um dos desafios mais importantes que os geneticistas enfrentam no século 21. Os traços complexos são de suprema importância na genética médica e agrícola. Para os seres humanos, a pressão arterial, o peso corporal, a suscetibilidade à depressão, os níveis séricos de colesterol e o risco de desenvolvimento de câncer ou outros distúrbios são, todos, traços complexos. Para as plantas de cultivo, o nível de produção, a resistência a patógenos, a capacidade de tolerar o estresse pela seca, a eficiência da captação de fertilizantes e até mesmo o sabor são, todos, traços complexos. Para o gado, a produção de leite em vacas leiteiras, a massa muscular no gado de corte, o tamanho da ninhada em porcos e a produção de ovos em galinhas são, todos, traços complexos. Apesar da importância dos referidos traços, sabemos muito menos a respeito de sua herança do que sabemos a respeito da herança de traços herdados de modo simples, tais como a fibrose cística ou a anemia falciforme.

Neste capítulo, exploraremos a herança dos traços complexos. Iniciaremos com uma revisão de alguns conceitos estatísticos básicos. Em seguida, desenvolveremos o modelo matemático utilizado para conectar a ação de genes na

célula com os fenótipos que observamos no nível do organismo inteiro. Com a utilização desse modelo, em seguida demonstraremos como os geneticistas quantitativos separam a variação fenotípica em uma população nas partes que ocorrem em virtude de fatores genéticos e ambientais. Revisaremos os métodos utilizados por cultivadores de plantas e criadores de animais para prever o fenótipo da descendência a partir do fenótipo de seus genitores. Finalmente, veremos como uma combinação da análise estatística e de marcadores moleculares pode ser utilizada para identificar os genes específicos que controlam os traços quantitativos.

19.1 Medida da variação quantitativa

Para estudar a herança de traços quantitativos, necessitamos de algumas ferramentas estatísticas básicas. Nesta seção, introduziremos a média, que pode ser utilizada para descrever as diferenças entre os grupos e a variância, que pode ser utilizada para mensurar a quantidade de variação que existe em um grupo. Também discutiremos sobre a distribuição normal, que é central para a compreensão da variação quantitativa nas populações. Mas antes de discutir as ferramentas estatísticas, definiremos os diferentes tipos de variação em traços complexos que podem ocorrer em uma população.

Tipos de traço e herança

Um **traço contínuo** é um traço que pode assumir um número possivelmente infinito de estados ao longo de uma variação contínua. A altura em seres humanos é um bom exemplo. As pessoas podem variar de aproximadamente 140 a 230 cm de altura. Se medíssemos a altura com precisão, o número possível de alturas seria infinito. Por exemplo, uma pessoa pode ter 170 cm de altura ou 170,2 ou 170,0002 cm. Os traços contínuos tipicamente apresentam uma **herança complexa**, que envolve múltiplos genes, além de fatores ambientais.

Em relação a alguns traços, os indivíduos em uma população podem ser classificados em grupos ou categorias distintas. Os referidos traços são conhecidos como **traços categóricos**. Exemplos incluem flores roxas *versus*

brancas ou caules altos *versus* baixos em relação às ervilhas de Mendel, conforme vimos no Capítulo 2. Os traços categóricos com frequência exibem uma **herança simples**, de tal modo que a progênie de cruzamentos segrega em razões mendelianas padrão, tais como 3:1 em relação a um gene único ou 15:1 em relação a dois genes. A herança é simples em virtude de apenas um ou dois genes estarem envolvidos e o ambiente apresentar pouco ou nenhum efeito sobre o fenótipo.

Alguns traços categóricos não demonstram herança simples. Esses incluem muitas condições de doenças em seres humanos. Em genética médica, os indivíduos podem ser classificados nas categorias "acometidos" ou "não acometidos" por uma doença. Por exemplo, um indivíduo pode ou não ter diabetes tipo 2. Entretanto, o diabetes tipo 2 não segue as regras mendelianas simples, nem produz razões mendelianas nos heredogramas. Em vez disso, existem múltiplos fatores genéticos e ambientais que impõem a uma pessoa o risco de desenvolvimento dessa doença. Os indivíduos que apresentam um determinado número de fatores de risco ultrapassarão um limiar e desenvolverão a doença. O diabetes tipo 2 é um traço categórico denominado **traço de limiar** e apresenta herança complexa.

Outro tipo de traço é o **traço merístico** ou traço de contagem, que assume uma diversidade de valores distintos. Um exemplo seria o tamanho da ninhada em aves. Uma ave pode pôr 1, 2, 3 ou mais ovos, mas não pode pôr 2,49 ovos. Os traços merísticos são quantitativos, mas estão restritos a valores definidos. Eles não assumem uma variação contínua de valores. Os traços merísticos normalmente apresentam herança complexa.

Geneticistas quantitativos buscam compreender a herança de traços que demonstram herança complexa, que resulta de uma combinação de fatores genéticos e ambientais. Eles podem investigar traços que são categóricos, merísticos ou contínuos. A ênfase está no tipo de herança – complexa. Por esse motivo, o termo *traço complexo* com frequência é preferido ao traço contínuo ou quantitativo, tendo em vista que inclui todos os tipos de traços com os quais a genética quantitativa se preocupa. Qualquer fenômeno biológico em relação ao qual existe uma variação pode demonstrar herança complexa e pode ser estudado

como um traço complexo. Portanto, o tamanho e a forma das estruturas, a cinética enzimática, os níveis de mRNA, os ritmos circadianos e os cantos das aves podem ser, todos, tratados como traços complexos.

Média

Quando os geneticistas quantitativos estudam a herança de um traço, eles trabalham com um grupo de indivíduos em particular, ou **população**. Por exemplo, podemos estar interessados na herança da altura em relação à população de homens adultos em Xangai, na China. Aqui, estamos utilizando "população" para indicar um grupo que compartilha determinadas características em comum, tais como idade, sexo, etnia ou origem geográfica. Tendo em vista que existem mais de 5 milhões de homens adultos em Xangai, a determinação de cada uma das suas alturas seria uma tarefa hercúlea. Portanto, geneticistas quantitativos estudam tipicamente apenas um subconjunto, ou **amostra**, da população total. A amostra deve ser escolhida *aleatoriamente*, de tal modo que cada um dos 5 milhões de homens apresente uma chance igual de ser incluído na amostra. Se a amostra satisfizer a esse critério, então podemos utilizar medidas realizadas na amostra para fazer inferências a respeito da população inteira.

Utilizando o exemplo da altura para os homens de Xangai, podemos descrever a população com a utilização da **média** ou do valor médio em relação ao traço. Selecionamos uma amostra aleatória de 100 homens da população e medimos as suas alturas. Alguns dos homens podem ter 166 cm de altura, outros 172 cm de altura e assim por diante. Para calcular a média, simplesmente somamos todas as medidas individuais e dividimos a soma pelo tamanho da amostra (n), que nesse caso é 100. Em relação aos dados na Tabela 19.1, o resultado seria 170 cm ou 5 pés e 7 polegadas. Tendo em vista que temos uma amostra aleatória, podemos inferir que a altura média da população inteira é de 170 cm.

A altura é uma *variável aleatória*, o que significa que pode assumir diferentes valores e, quando selecionamos alguém da população aleatoriamente, o valor que observamos é regulado por um elemento ao acaso. As variáveis aleatórias normalmente são representadas pela letra X em estatística. Temos as medidas de X_1, X_2, X_3, … X_{100} em relação aos n = 100 homens na amostra. Simbolicamente,

podemos expressar a média como:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

em que $\overline{X}$ representa a média da amostra. A letra grega maiúscula sigma (Σ) é o sinal de somatório, que indica que adicionamos todos os *n* valores de *X* observados em relação a *i* = 1, 2, até *n*. (Com frequência, o *n* sobre Σ e o *i* = 1 abaixo de Σ são omitidos para simplificar o aspecto das equações.)

Existe uma distinção que é feita entre a média de uma amostra $\bar{x}$ e a média verdadeira da população. Para conhecer a média verdadeira em relação à altura dos homens em Xangai, precisaríamos determinar a altura de todo e cada homem. A média verdadeira é simbolizada pela letra grega μ e, assim, temos diferentes símbolos para as médias da amostra e da população.

Aqui está outro modo para calcular a média, que com frequência é consideravelmente útil. Podemos adicionar os produtos de cada classe de valores de *X* no conjunto de dados vezes a frequência daquela classe no conjunto de dados. Essa operação é simbolizada como:

$$\overline{X} = \sum_{i=1}^{k} f_i X_i$$

em que f_i é a frequência da *i*-ésima classe de observações, X_i é o valor da *i*-ésima classe e existe um total de *k* classes. Em relação aos dados na Tabela 19.1, um homem dos 100 (*f* = 0,01) tem 156 cm de altura, dois homens (*f* = 0,02) têm 157 cm de altura, e assim por diante e, portanto, podemos calcular a média da amostra como:

$\overline{X}$ = (0,01 × 156) + (0,02 × 157) + … + (0,02 × 184) = 170

A média é útil para descrever populações e para comparar as diferenças entre as populações. Por exemplo, homens em áreas urbanas da China têm em média 170 cm de altura, enquanto homens em áreas rurais da China têm 166 cm de altura. Esses valores foram calculados com a utilização de amostras coletadas de cada

região. Uma questão que um geneticista quantitativo pode indagar a respeito da diferença observada na altura entre homens chineses rurais e urbanos é a seguinte: “A diferença ocorre em virtude de fatores genéticos ou ocorre em virtude de diferenças na nutrição, nos cuidados de saúde ou outros fatores ambientais?” Posteriormente no capítulo, veremos como os geneticistas quantitativos desmembram as contribuições genéticas *versus* ambientais para um traço.

Por último, existe outra anotação útil da estatística que pode ser utilizada para definir a média. A média de uma variável aleatória, *X*, é a *expectativa* ou o *valor esperado* daquela variável aleatória. O valor esperado é a média de todos os valores que observaríamos se medíssemos *X* muitas vezes. A expectativa é simbolizada por *E*, e escrevemos *E(X)* para significar “o valor esperado de *X*”. Simbolicamente, escrevemos:

$$E(X) = \overline{X}$$

Tabela 19.1 Dados simulados em relação à altura de 100 homens de Xangai, na China.

Altura (cm)	Contagem	Frequência × Altura
156	1	1,56
157	2	3,14
158	1	1,58
159	2	3,18
160	1	1,60

161	1	1,61
162	2	3,24
164	7	11,48
165	7	11,55
166	1	1,66
167	6	10,02
168	9	15,12
169	7	11,83
170	9	15,30
171	5	8,55
172	5	8,60
173	6	10,38
174	5	8,70
175	6	10,50

176	3	5,28
177	4	7,08
178	2	3,56
179	2	3,58
180	2	3,60
181	2	3,62
184	2	3,68
Soma	100	170,0

Utilizaremos a anotação de expectativa em diversos locais neste capítulo.

Variância

Além da média, também precisamos de uma medida para quantificar a variação existente nas populações. Podemos criar uma representação visual da variação ao inserir em um gráfico a contagem ou a frequência de cada classe de altura. A Figura 19.1 demonstra um referido gráfico em relação aos nossos dados de altura simulados para 100 homens de Xangai. O eixo *x* demonstra diferentes classes de altura e o eixo *y* demonstra a contagem ou a frequência de cada classe. Nessa figura, os homens foram acondicionados em grupos de 4 cm, por exemplo, entre 155 e 158 cm. Esse tipo de gráfico é denominado **histograma de frequência**. Se os valores estiverem agrupados firmemente ao redor da média, existe então menos

variação, e se os valores estiverem dispersos ao longo do eixo *x*, existe maior variação.

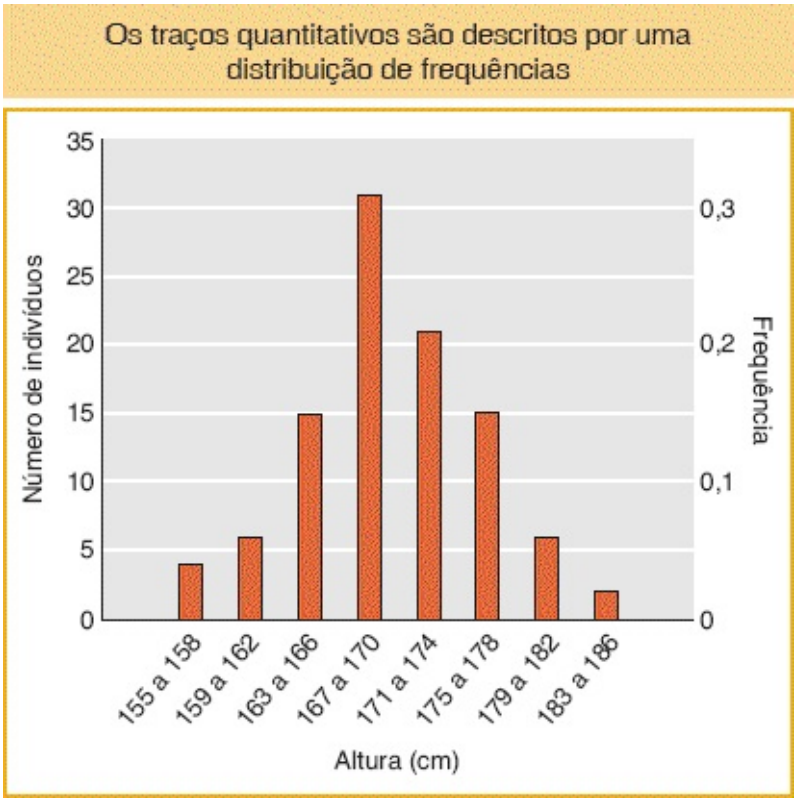


FIGURA 19.1 Histograma de frequências de dados simulados em relação à altura de homens adultos de Xangai, na China.

Podemos quantificar a variação em uma população ao utilizar uma medida estatística denominada **variância**. A variância mede a extensão até a qual os indivíduos na população se desviam da média da população. Se todos os 100 homens em nossa amostra apresentassem alturas muito próximas da média, a variância seria então pequena. Se as suas alturas se desviassem amplamente da média, a variância seria grande.

Tendo em vista que a variância é uma medida de **desvio** da média, definiremos o desvio matematicamente. Conhecendo o valor médio em relação à variável aleatória *X*, podemos calcular o desvio do referido indivíduo da média ao subtrair $\overline{X}$ das observações individuais. Representaremos os desvios por um *x* minúsculo:

$$x = X - \bar{X}$$

Alguns indivíduos apresentarão valores de X acima da média e eles apresentarão um desvio positivo. Outros apresentarão valores de X abaixo da média e eles apresentarão um desvio negativo. Em relação à população em geral, o valor esperado de x é 0 ou $E(x) = 0$.

Para medir a quantidade de variação em relação a X na população, utilizamos a variância, que é a média dos desvios ao quadrado. Primeiramente, calculamos a soma dos desvios ao quadrado (ou *soma dos quadrados*, abreviadamente) como:

$$\begin{aligned}\text{Soma dos quadrados} &= \sum_i (X_i - \bar{X})^2 \\ &= \sum_i (x_i)^2\end{aligned}$$

Tendo em vista que os desvios com valores negativos formam quadrados positivos, ambos os desvios negativos e positivos contribuirão positivamente para a soma dos quadrados. A variância é a média dos desvios ao quadrado (ou a soma dos quadrados dividida por n). Simbolicamente, expressamos a variância da população como:

$$\begin{aligned}V_X &= \frac{1}{n}\sum_i (X_i - \bar{X})^2 \\ &= \frac{1}{n}\sum_i (x_i)^2\end{aligned}$$

em que V_X indica a variância de X. A variância da população por vezes é simbolizada com a utilização da letra grega sigma minúscula ao quadrado (σ^2). Em estatística, também é feita uma distinção entre a variância da população (σ^2) e a variância da amostra (s^2). A última é calculada ao dividir as somas dos quadrados por n — 1, em vez de n, para corrigir um desvio causado por um tamanho de amostra pequeno. Para simplicidade, utilizaremos a variância da população e a fórmula anterior em todo este capítulo.

Existem diversos pontos a serem compreendidos a respeito da variância.

Primeiramente, a variância fornece uma medida da dispersão ao redor da média. Quando a variância é alta, os valores indivíduos estão dispersos até mais distante da média; quando ela é baixa, então os valores individuais são agrupados mais próximo da média. Em segundo lugar, a variância é medida em unidades ao quadrado, de tal modo que se medirmos a altura humana em centímetros, a variância então será em centímetros². Em terceiro lugar, a variância pode variar de 0,0 até o infinito. Em quarto lugar, a variância é igual ao valor esperado do desvio elevado ao quadrado (x^2) ou $E(x^2)$.

A variância dos traços quantitativos é medida em unidades ao quadrado. Essas unidades ao quadrado apresentam propriedades matemáticas desejáveis, conforme veremos a seguir; entretanto, elas não fazem sentido intuitivamente. Se medirmos o peso em quilogramas, a variância então será em quilogramas², que não apresenta um sentido claro. Portanto, outra estatística utilizada para quantificar a extensão do desvio da média em uma população é o **desvio padrão** (σ), que é a raiz quadrada da variância:

$$\sigma = \sqrt{\sigma^2}$$

O desvio padrão é expresso nas mesmas unidades que o próprio traço e, assim, o seu significado é mais intuitivo. Utilizaremos o desvio padrão na descrição dos traços a seguir.

Distribuição normal

Até mesmo se você nunca fez um curso de estatística, provavelmente você ouviu falar da **distribuição normal**, também conhecida como a "curva sinosoidal". A distribuição normal é notavelmente útil na biologia em geral e na genética quantitativa em particular, tendo em vista que a distribuição de frequência em relação a muitos traços biológicos se aproxima de uma curva normal. Por esse motivo, geneticistas podem tirar vantagem de diversas características da distribuição normal para descrever os traços quantitativos e dissecar a genética subjacente.

A distribuição normal é uma distribuição de frequência contínua semelhante ao

histograma de frequência demonstrado na Figura 19.1. A distribuição normal se aplica aos traços contínuos. Conforme mencionado anteriormente, os traços contínuos podem assumir um número infinito de valores. Uma pessoa pode ter 170 cm de altura ou 170,2 ou 170,002 e assim por diante. Em relação a tais traços, a frequência esperada de diferentes valores de traço é mais bem-representada por uma curva do que por um histograma de frequência. Em relação à distribuição normal, o formato da curva é determinado por dois fatores – a média e o desvio padrão.

Aqui está um exemplo com a utilização de dados da altura em relação a 660 mulheres dos EUA, coletados pelos Centers for Disease Control and Prevention. O histograma de frequência demonstra o formato clássico da curva sinusoidal com o pico próximo do valor médio de 164,4 cm e os valores fora da média distribuídos simetricamente ao redor da média (Figura 19.2 A). Podemos adaptar uma curva normal a essa distribuição utilizando apenas duas informações – a média e o desvio padrão. A forma da curva é definida por uma equação denominada função de densidade da probabilidade normal, na qual a média e o desvio padrão estão conectados. A distribuição normal possibilita que prevejamos a porcentagem das observações que estará situada dentro de uma determinada distância da média (Figura 19.2 B). Se medirmos a distância ao longo do eixo *x* em desvios padrões, então é esperado que 68% das observações estejam situadas dentro de 1 desvio padrão (σ) da média e 95,5% dentro de 2 desvios padrões. Em relação aos dados da altura das mulheres norte-americanas, 71% (449 mulheres) estão situadas dentro de 1 desvio padrão da média e 96% (633 mulheres) dentro de 2 desvios padrões. Esses valores estão muito próximos das previsões de 68,2% e 95,5% com base na curva normal.

Se conhecermos apenas a média e o desvio padrão em relação a um traço, podemos prever a forma da distribuição do traço na população e podemos prever como provavelmente observaremos determinados valores na realização de amostragem da população. Por exemplo, se a altura média em relação às mulheres norte-americanas é de 164,4 cm (5 pés e 5 polegadas) e o desvio padrão é de 6,18 cm, podemos prever que apenas 2% das mulheres terão mais de 177 cm de altura ou 5 pés e 10 polegadas. Conforme demonstrado na Figura 19.2 C, se o

desvio padrão for maior (p. ex., 8), então a curva será mais achatada e maior porcentagem estará situada acima de 177 cm. Entretanto, ainda seria verdadeiro que apenas 2% estariam a mais de 2σ acima da média ou 180,4 cm [(164,4 + (2 × 8)].

> CONCEITO-CHAVE Um traço complexo é qualquer traço que não demonstre herança mendeliana simples. Um traço complexo pode ser um traço descontínuo, tal como a presença ou a ausência de uma condição de doença, ou um traço continuamente variável, tal como a altura em seres humanos. O campo da genética quantitativa estuda a herança de traços complexos com a utilização de algumas ferramentas estatísticas básicas, incluindo a média, a variância e a distribuição normal.

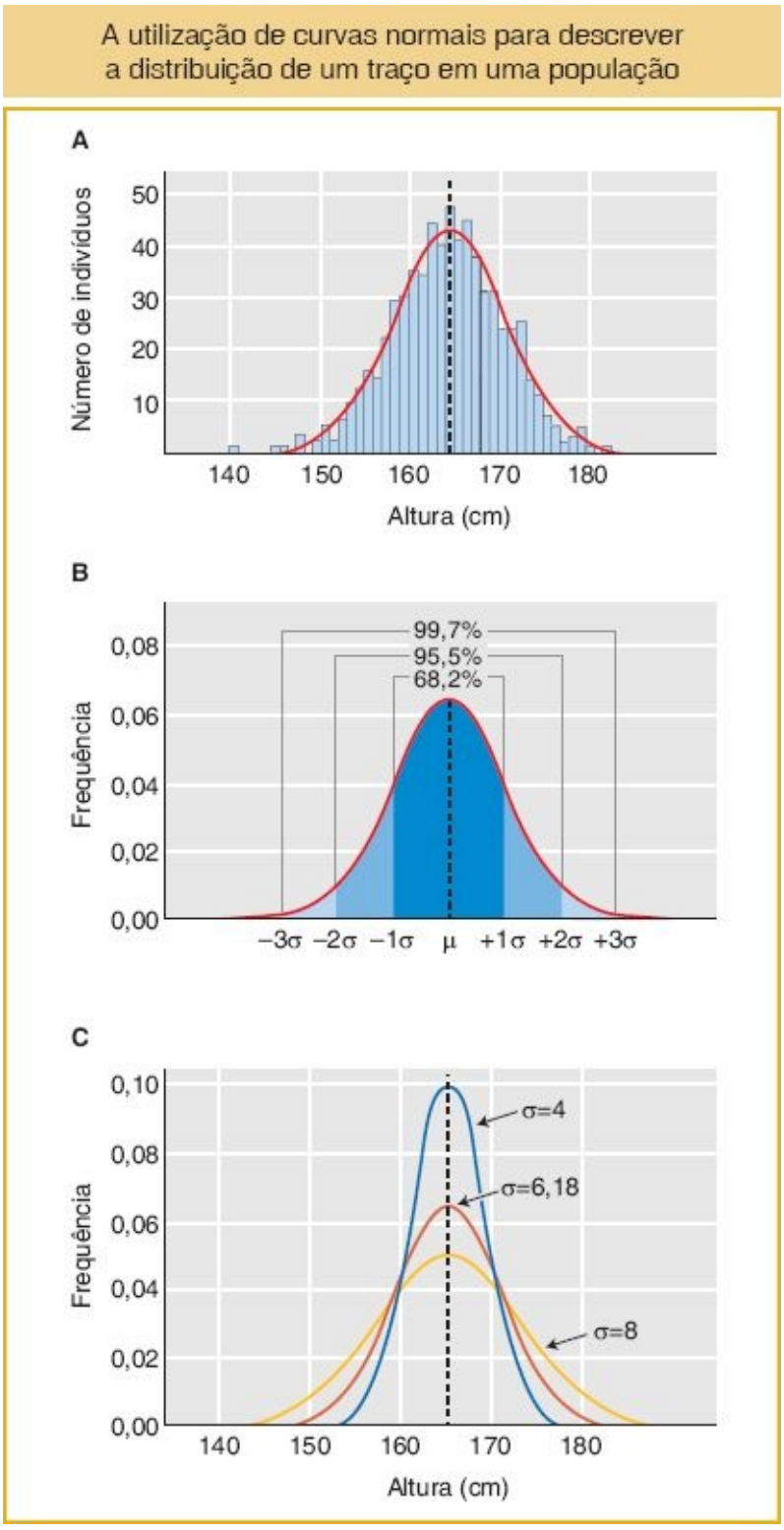


FIGURA 19.2 A. Histograma de frequência de dados reais em relação à altura de mulheres adultas dos EUA. A *linha vermelha* representa a adaptação da curva normal a estes dados com uma média de 164,4 cm e desvio padrão de 6,18 cm. **B.** Curva normal em relação à altura de mulheres norte-americanas,

demonstrando as porcentagens previstas de mulheres que estarão classificadas dentro de diferentes números de desvios padrões da média. C. Curvas normais com a mesma média (164,4 cm), mas diferentes desvios padrões, demonstrando o efeito do desvio padrão sobre a forma da curva.

19.2 Modelo genético simples para os traços quantitativos

Um modelo matemático é uma representação simplificada de um fenômeno complexo. Como um exemplo, podemos colocar um balde sob uma torneira aberta e medir o volume de água no balde como uma função da quantidade de tempo que o balde é deixado sob a torneira: *volume = função (tempo)*. Podemos construir um modelo mais detalhado, que inclua a velocidade na qual a água sai da torneira: *volume = função (velocidade × tempo)*. Os modelos nos possibilitam descrever um fenômeno em termos das variáveis que o influenciam e em seguida utilizar o modelo para fazer previsões a respeito do estado do fenômeno sob valores diferentes em relação a essas variáveis. Nesta seção, definiremos o modelo matemático utilizado por geneticistas quantitativos para estudar traços complexos.

Desvios genéticos e ambientais

Agora examinaremos como os fenótipos podem ser decompostos em suas contribuições genéticas e ambientais, utilizando como um exemplo a altura de Yao Ming, o ex-pivô do time de basquete Houston Rockets. Yao Ming tem a altura de 229 cm ou 7 pés e 6 polegadas (Figura 19.3). Isso mesmo: Yao Ming é quase 60 cm mais alto do que o homem médio de Xangai, que é justamente a cidade natal de Yao Ming. Assim como em relação a todos nós, a altura de Yao Ming é o resultado combinado do seu genótipo e do ambiente no qual ele cresceu. Faremos um experimento imaginário e veremos como podemos desmembrar as contribuições genéticas e ambientais em relação à altura excepcional de Yao Ming.

Primeiramente, definiremos um modelo matemático simples, que possa ser aplicado para qualquer traço quantitativo. O valor para um indivíduo em relação

a um traço (X) pode ser expresso em termos da média da população e dos desvios da média em virtude de fatores genéticos (g) e ambientais (e).

$$X = \overline{X} + g + e$$

Estamos utilizando as letras minúsculas g e e para os desvios genéticos e ambientais, assim como utilizamos uma letra minúscula x para o desvio de X da média. Portanto, no caso de Yao Ming, a sua altura pode ser expressa como o valor médio dos homens de Xangai (170 cm), mais os seus desvios genético e ambiental específicos ($g + e$ = 59 cm). Podemos simplificar a equação anterior ao subtrair $\overline{X}$ de ambos os lados para obter:

$x = g + e$

A altura excepcional de Yao Ming

FIGURA 19.3 O astro pivô de basquete Yao Ming, que tem 229 cm ou 7 pés e 6 polegadas de altura, passando por um grupo de praticantes de marcha atlética. (STR/AFP/Getty Images/Newscom.)

em que x representa o desvio fenotípico do indivíduo. Para Yao Ming, $x = g + e$ = 59.

Como podemos determinar os valores de g e e em relação a Yao Ming? Um modo seria se tivéssemos clones de Yao Ming (os clones são indivíduos geneticamente idênticos). Imaginemos que clonamos Yao Ming e que distribuímos

esses clones (quando recém-nascidos) para um grupo de famílias escolhidas aleatoriamente em Xangai. Vinte e um anos depois, localizamos esse exército de clones de Yao Ming, medimos as suas alturas e determinamos que a sua altura média é de 212 cm. A expectativa de *e* nos muitos ambientes nos quais os clones de Yao Ming foram criados é 0. Em algumas famílias, os clones obtêm um ambiente positivo (+*e*) e em outras, um ambiente negativo (–*e*). Em geral, *E(e)* = 0. Portanto, a média para os clones menos a média da população é igual ao desvio genotípico de Yao Ming ou *g* = (212 – 170) = 42 cm. Os 17 cm remanescentes de seu extraordinário desvio fenotípico de 59 cm é *e* para o ambiente específico no qual o Yao Ming real cresceu. Acoplando esses valores na equação anterior, obtemos:

$$229 = 170 + 42 + 17$$

Concluímos que a altura excepcional de Yao Ming ocorre principalmente em virtude da genética excepcional, mas ele também vivenciou um ambiente que reforçou a sua altura.

Embora o nosso experimento imaginário de clonagem de Yao Ming seja exagerado, muitas espécies de plantas e algumas espécies de animais podem ser propagadas pela clonagem com facilidade. Por exemplo, pode-se utilizar “cortes” de uma planta individual para produzir múltiplos indivíduos geneticamente idênticos. Outro modo de criar indivíduos geneticamente idênticos é por meio da produção de **linhagens** ou **cepas endocruzadas** (Quadro 19.1). Todos os indivíduos nas referidas linhagens são geneticamente idênticos, tendo em vista que são totalmente endocruzados a partir de um genitor ou genitores em comum. Ao utilizar clones ou linhagens endocruzadas, os geneticistas podem estimar as contribuições genéticas e ambientais em relação a um traço ao criar os clones em ambientes atribuídos aleatoriamente. Aqui está um exemplo.

A Tabela 19.2 (experimento I) demonstra dados simulados em relação a 10 linhagens endocruzadas de milho que foram cultivadas em três ambientes diferentes e pontuadas em relação ao número de dias entre o plantio e a ocasião em que as plantas liberaram pólen pela primeira vez. A média geral é de 70 dias.

Consideraremos a linhagem A quando cultivada no ambiente 1. A média para todas as linhagens no ambiente 1 é de 68 ou 2 a menos do que a média geral e, assim, *e* para o ambiente 1 é –2. A média da linhagem A em todos os três ambientes é de 64 ou 6 a menos do que a média geral e, assim, *g* para a linhagem A é –6. Reunindo esses dois valores, decompomos o fenótipo da linhagem A quando cultivada no ambiente 1 como:

Quadro 19.1 Linhagens endocruzadas.

Uma **linhagem endocruzada** é uma linhagem específica de uma espécie de planta ou animal que foi autofertilizada ou submetida ao cruzamento entre irmãos durante múltiplas gerações, de tal modo que ela se torna homozigota (ou endocruzada) na maior parte de seu genoma. A autofertilização pode ser utilizada em espécies hermafroditas, tais como a maior parte das plantas. Nesse processo, apenas uma semente é utilizada para formar cada geração subsequente. No milho, por exemplo, uma única planta individual é escolhida e autopolinizada. Em seguida, na próxima geração, uma única planta de sua descendência é escolhida e autopolinizada. Na terceira geração, uma única planta de sua descendência é escolhida e autopolinizada, e assim por diante. Suponha que a planta original seja uma heterozigota (*A*/*a*); a autopolinização produzirá então descendência que é $\frac{1}{2}$ heterozigota e $\frac{1}{2}$ homozigota ($\frac{1}{4}$ *A*/*A* mais $\frac{1}{4}$ *a*/*a*). Do conjunto de todos os *loci* heterozigotos no genoma, então, após uma geração de autopolinização, apenas $\frac{1}{2}$ ainda será heterozigota; após duas gerações, $\frac{1}{4}$; após três gerações, $\frac{1}{8}$ e assim por diante. Na *n*-ésima geração:

$$\text{Het}_n = \frac{1}{2^n}\text{Het}_0$$

em que Het_n é a proporção de *loci* heterozigotos na *n*-ésima geração e Het_0 é a proporção na geração 0.

Quando a autofecundação não é possível, o cruzamento entre irmão e irmã alcançará a mesma finalidade, embora mais lentamente. A tabela demonstra a quantidade de heterozigosidade remanescente após *n* gerações de

autofecundação e cruzamento entre irmão e irmã.

Heterozigosidade remanescente

Geração	Autofecundação	Cruzamento irmão-irmã
0	1,000	1,000
1	0,500	0,750
2	0,250	0,625
3	0,125	0,500
4	0,0625	0,406
5	0,03125	0,338
10	0,000977	0,114
20	$0{,}95 \times 10^{-6}$	0,014
n	$\mathrm{Het}_n = \frac{1}{2}\mathrm{Het}_{n-1}$	$\mathrm{Het}_n = \frac{1}{2}\mathrm{Het}_{n-1} + \frac{1}{4}\ \mathrm{Het}_{n-2}$

As linhagens endocruzadas são imensamente importantes não apenas na genética quantitativa, mas na genética em geral. Geneticistas desenvolveram muitas linhagens endocruzadas para diferentes organismos-modelo, incluindo *Drosophila*, camundongo, *C. elegans*, levedura, *Arabidopsis* e milho. Se uma linhagem endocruzada for utilizada em um experimento, sabe-se então que os indivíduos que recebem diferentes tratamentos são geneticamente idênticos. Portanto, quaisquer diferenças observadas entre os tratamentos não podem ser atribuídas às diferenças genéticas entre os indivíduos utilizados no experimento.

Tabela 19.2 Dados simulados em relação aos dias até a polinização para 10 linhagens endocruzadas de milho cultivadas em dois experimentos.

Experimento I										
Linhagens endocruzadas	A	B	C	D	E	F	G	H	I	J
Ambiente 1	62	64	66	66	68	68	70	70	72	7
Ambiente 2	64	66	68	68	70	70	72	72	74	7
Ambiente 3	66	68	70	70	72	72	74	74	76	7
Média	64	66	68	68	70	70	72	72	74	7
Experimento II										
Linhagens endocruzadas	A	B	C	D	E	F	G	H	I	J
Ambiente 4	58	60	62	62	64	64	66	66	68	6
Ambiente 5	64	66	68	68	70	70	72	72	74	7
Ambiente 6	70	72	74	74	76	76	78	78	80	8
Média	64	66	68	68	70	70	72	72	74	7

$$62 = 70 + (–6) + (–2)$$

Podemos fazer os mesmos cálculos em relação às outras nove linhagens endocruzadas e em seguida teríamos uma descrição completa de todos os fenótipos em cada ambiente em termos da extensão na qual o seu desvio da média geral ocorre em virtude de fatores genéticos e ambientais.

Variâncias genética e ambiental

Podemos utilizar o modelo simples $x = g + e$ para pensar adicionalmente a respeito da variância de traços quantitativos. Relembre que a variância é um modo de medir o quanto os indivíduos se desviam da média da população. Sob esse modelo, a variância do traço pode ser separada nas variâncias genética e ambiental:

$$V_X = V_g + V_e$$

Esta equação simples nos informa que o traço ou a variação fenotípica (V_X) é a soma de dois componentes – a variância genética (V_g) e a variância ambiental (V_e). Conforme observado no Quadro 19.2, existe uma presunção importante por trás dessa equação; a saber, que o genótipo e o ambiente não estão correlacionados – ou seja, eles são independentes. Se os melhores genótipos estão inseridos nos melhores ambientes e os piores genótipos nos piores ambientes, essa equação então fornece resultados imprecisos. Discutiremos essa presunção importante posteriormente no capítulo.

Podemos utilizar os dados da Tabela 19.2 (experimento I) para explorar a equação em relação às variâncias. Primeiramente, utilizaremos todos os 30 valores fenotípicos em relação às 10 linhagens nos três ambientes para calcular a variância. O resultado é V_X = 14,67 dias². Agora, para estimar V_g, calculamos a variância das médias entre as 10 linhagens endocruzadas. O resultado é V_g = 12,0 dias². Finalmente, para estimar V_e, calculamos a variância das médias entre os três ambientes. O resultado é V_e = 2,67 dias². Portanto, a variância fenotípica

(14,67) é igual à variância genética (12,0) mais a variância ambiental (2,67). A equação funciona em relação a esses dados porque o genótipo e o ambiente não estão correlacionados. Todos os genótipos experimentam a mesma variação de ambientes.

Se calcularmos os desvios padrões em relação aos dados na Tabela 19.2 (experimento I), observaremos que o desvio padrão fenotípico (3,83) não é a soma dos desvios padrões genético (3,46) e ambiental (1,63). As variâncias podem ser decompostas em diferentes fontes. Os desvios padrões não podem ser decompostos dessa maneira. A seguir, veremos como essa propriedade da variância é útil para quantificar a extensão na qual a variação do traço é hereditária *versus* ambiental.

Quadro 19.2 Variâncias genética e ambiental.

Para compreender melhor a equação básica $V_X = V_g + V_e$, precisamos introduzir um novo conceito de estatística – a **covariância**. A covariância proporciona uma medida da associação entre os traços. Em relação a duas variáveis aleatórias *X* e *Y*, a sua covariância é:

$$COV_{X,Y} = \frac{1}{n}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$$
$$= \frac{1}{n}\sum_i (x_i y_i)$$

em que *x* e *y* são os desvios de *X* e *Y* de suas respectivas médias, conforme descrito no texto principal. O termo $(X_i - \bar{X})(Y_i - \bar{Y})$ ou $(x_i y_i)$ é denominado *produto cruzado*. A covariância é obtida somando todos os produtos cruzados em conjunto e dividindo por *n*. A covariância é a média ou o valor esperado, $E(xy)$, dos produtos cruzados. A covariância pode variar de infinito negativo até infinito positivo. Se valores grandes de *X* estiverem associados a valores grandes de *Y*, a covariância então será positiva. Se valores grandes de *X* estiverem associados a valores pequenos de *Y*, então a covariância será negativa. Se não houver associação entre *X* e *Y*, a covariância então será zero.

Em relação aos traços independentes, a covariância será zero.

No texto principal, vimos que a variância é o valor esperado dos desvios ao quadrado:

$$V_X = E(x^2)$$

Tendo em vista que o desvio fenotípico (x) é a soma dos desvios genotípicos (g) e ambientais (e), podemos substituir ($g + e$) por x e obter:

$$\begin{aligned} \mathrm{Vx} &= E[(g + e)^2] \\ &= E[g^2 + e^2 + 2ge] \\ &= E(g^2) + \mathrm{E}(e^2) + \mathrm{E}(2ge) \end{aligned}$$

O primeiro termo [$E(g^2)$] é a **variância genética**, o termo intermediário [$E(e^2)$] é a **variância ambiental** e o último termo é o dobro da covariância entre o genótipo e o ambiente.

Em experimentos controlados, os diferentes genótipos são inseridos em diferentes ambientes aleatoriamente. Em outras palavras, o genótipo e o ambiente são independentes. Se o genótipo e o ambiente forem independentes, então a covariância entre o genótipo e o ambiente $E(ge) = 0$ e a equação é reduzida para:

$$\begin{aligned} V_x &= E(g^2) + E(e^2) \\ &= V_g + Ve \end{aligned}$$

Portanto, a variância fenotípica é a soma da variância em virtude dos diferentes genótipos na população e da variância em virtude dos diferentes ambientes nos quais os organismos são criados.

Finalmente, veremos o que ocorre com as variâncias se o genótipo e o ambiente estiverem correlacionados. Para tanto, imagine que conhecemos os desvios genéticos (g) de nove cavalos puros-sangues-ingleses em relação ao

tempo necessário para que eles corram no Kentucky Derby. Também conhecemos os desvios ambientais (*e*) com os quais seus treinadores contribuem para que cada cavalo dispute essa corrida. Suporemos que, além do treinamento, não existem outras fontes de variação ambiental. A média da população em relação a esse conjunto de puros-sangues-ingleses é de 123 segundos para correr no Derby. Atribuímos os melhores cavalos aos melhores treinadores e os piores cavalos aos piores treinadores. Ao fazer isso, criamos uma relação não aleatória ou uma correlação entre os cavalos (genótipos) e os treinadores (ambientes).

A Tabela 19.3 demonstra os dados em relação a esse experimento imaginário. Você observará que V_X (6,67) não é igual à soma de V_g (2,22) e V_e (1,33). Tendo em vista que o genótipo e o ambiente estão correlacionados, violamos a presunção da equação que declara que $V_X = V_g + V_e$. A equação funciona apenas quando o genótipo e o ambiente não são correlacionados.

Correlação entre variáveis

Se o genótipo e o ambiente estiverem correlacionados, então a equação $V_X = V_g + V_e$ não será aplicável. Em vez disso, para que essa equação seja apropriada, o genótipo e o ambiente devem ser não correlacionados ou independentes. Vejamos um pouco mais de perto o conceito de **correlação**, a existência de uma relação entre duas variáveis. Esse é um conceito crítico para a genética quantitativa, conforme veremos neste capítulo.

Para visualizar o grau de correlação entre duas variáveis, podemos construir gráficos ou diagramas de dispersão. A Figura 19.4 demonstra os gráficos de dispersão que observaríamos sob diversas forças de correlação diferentes entre duas variáveis. Esses gráficos utilizam dados simulados em relação às alturas de conjuntos imaginários de gêmeos adultos idênticos. O painel superior da figura demonstra uma correlação perfeita, que é o que observaríamos se a altura de um gêmeo fosse exatamente a mesma do outro gêmeo para todos os conjuntos de gêmeos. O painel intermediário demonstra uma correlação forte, porém não perfeita. Aqui, quando um gêmeo é baixo, o outro também tende a ser baixo e quando um é alto, o outro tende a ser alto. O painel inferior demonstra a relação que observaríamos se a altura de um gêmeo não estivesse correlacionada com a

do outro gêmeo do conjunto. Aqui, a altura de um gêmeo de cada conjunto é aleatória a respeito do outro gêmeo do conjunto. Na próxima seção, veremos que os dados em relação a gêmeos reais se assemelhariam a algo como o painel intermediário.

Em estatística, existe uma medida específica de correlação, denominada **coeficiente de correlação**, que é simbolizada por uma letra minúscula *r*. É uma medida da associação entre duas variáveis. O coeficiente de correlação está relacionado com a covariância, que foi introduzida no Quadro 19.2; entretanto, ele é dimensionado para variar entre –1 e +1. Se simbolizarmos uma variável aleatória por *X* e a outra por *Y*, então o coeficiente de correlação entre *X* e *Y* é:

$$r_{X,Y} = \frac{COV_{X,Y}}{\sqrt{V_X V_Y}}$$

O termo $\sqrt{V_X V_Y}$ é utilizado para dimensionar a covariância para variar entre –1 e +1. A equação expandida em relação ao coeficiente de correlação é:

$$r_{X,Y} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \Sigma(Y_i - \bar{Y})^2}}$$

A equação é trabalhosa e, na prática, o cálculo do coeficiente de correlação é realizado com o auxílio de computadores. Em relação a duas variáveis que estão perfeitamente correlacionadas, $r = +1{,}0$ se, na medida em que uma variável aumentar, a outra aumentar, ou $r = -1{,}0$ se, na medida em que uma aumentar, a outra diminuir. Em relação a variáveis completamente independentes, $r = 0{,}0$.

Tabela 19.3 Dados simulados em relação ao tempo em segundos (*X*) de corrida dos cavalos no Kentucky Derby, decomposto em desvios genético (*g*) e ambiental (*e*) da média da população.

Cavalo	Média da população	*g*	Treinador	*e*	*x*	*X*

Secretariat	123	–2	Lucien	–2	–4	119
Decidedly	123	–2	Horatio	–1	–3	120
Barbaro	123	–1	Mike	–1	–2	121
Unbridled	123	–1	Carl	0	–1	122
Ferdinand	123	0	Charlie	0	0	123
Cavalcade	123	1	Bob	0	1	124
Meridian	123	1	Albert	1	2	125
Whiskery	123	2	Fred	1	3	126
Gallant Fox	123	2	Jim	2	4	127
Média (s)	123	0		0	0	123
Variância (s^2)		2,22		1,33	6,67	6,67

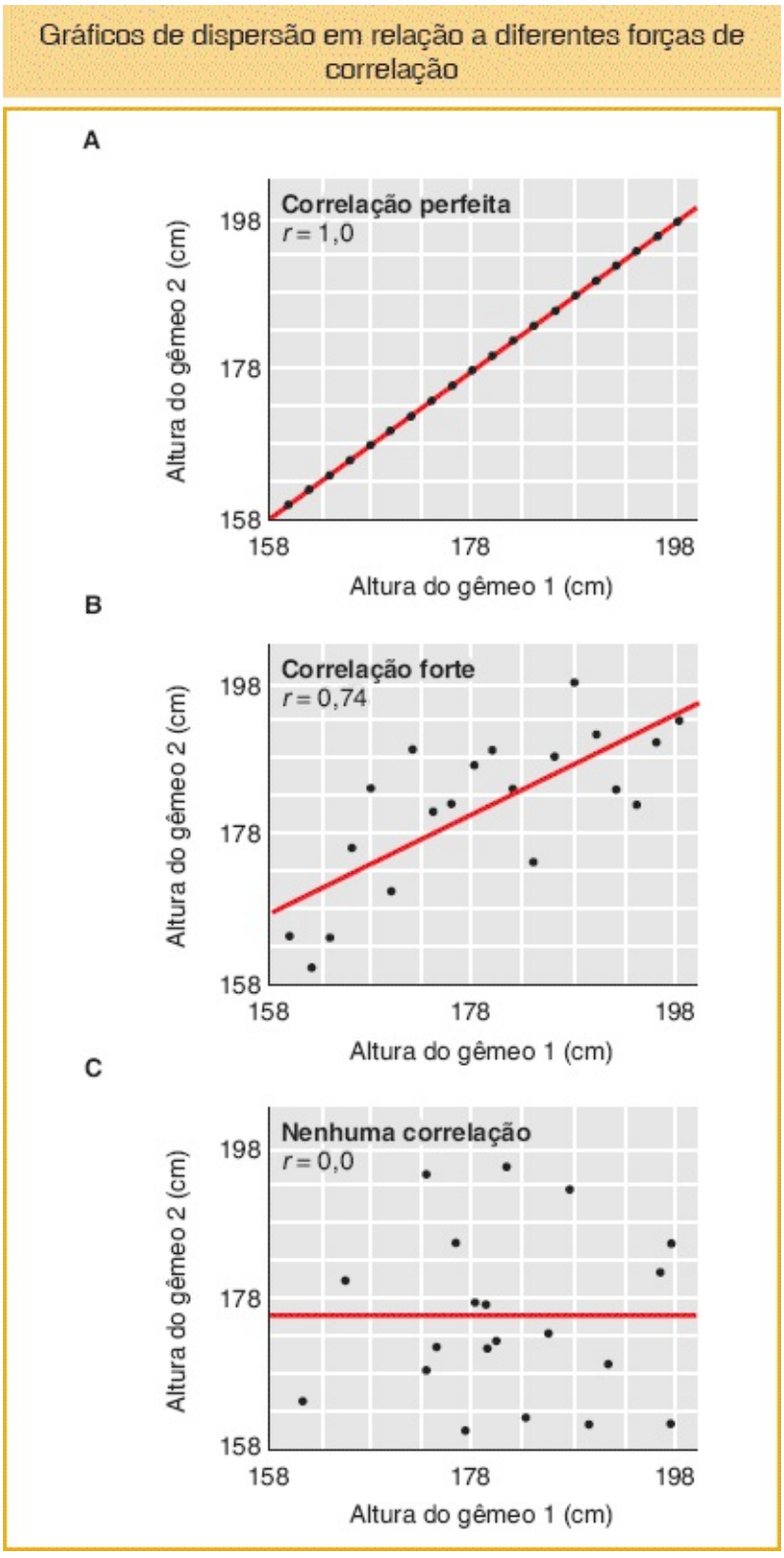


FIGURA 19.4 Gráficos de dispersão para o caso de uma correlação perfeita (**A**), uma correlação forte (**B**) e nenhuma correlação (**C**). As *linhas vermelhas* apresentam uma inclinação igual ao coeficiente de correlação.

Na Figura 19.4, o coeficiente de correlação é demonstrado em cada painel. Ele é 1,0 no painel superior para a uma correlação positiva perfeita, 0,74 no painel intermediário para uma correlação forte e 0,0 no painel inferior para nenhuma correlação (independência de *X* e *Y*). A inclinação da linha vermelha em cada painel é igual ao coeficiente de correlação e fornece um indicador visual da força da correlação.

Como um exercício, utilize os dados na Tabela 19.3 para construir um diagrama de dispersão e calcular o coeficiente de correlação. Isso seria mais bem-realizado com um computador e um *software* de planilhas. Utilize os desvios genéticos (*g*) no eixo *x* e os desvios ambientais (*e*) no eixo *y*. Em seguida calcule o coeficiente de correlação entre *g* e *e*. O diagrama de dispersão será semelhante àquele na Figura 19.4 (painel intermediário) e o coeficiente de correlação será 0,90. Portanto, quando os melhores cavalos são colocados com os melhores treinadores, a genética e o ambiente estão correlacionados e o modelo $V_X = V_g + V_e$ não pode ser utilizado.

> CONCEITO-CHAVE O fenótipo de um indivíduo em relação a um traço pode ser expresso em termos do seu desvio da média da população. O desvio fenotípico (*x*) de um indivíduo é composto por duas partes – seu desvio genético (*g*) e seu desvio ambiental (*e*). Experimentos com clones ou linhagens endocruzadas podem ser utilizados para decompor o fenótipo de um indivíduo em seus componentes genético e ambiental.
>
> A variação fenotípica em uma população em relação a um traço (V_X) pode ser decomposta nas variâncias genética (V_g) e ambiental (V_e). Essa decomposição presume que os genótipos e os ambientes não são correlacionados.

19.3 Herdabilidade no sentido amplo | Natureza versus criação

Uma questão-chave em genética é: “Quanto da variação em uma população ocorre em virtude de fatores genéticos e quanto em virtude de fatores ambientais?” Na imprensa popular, essa questão com frequência é expressa em termos de *natureza* versus *criação* – ou seja, qual é a influência dos fatores inatos (genéticos) em comparação com os fatores externos (ambientais)? As respostas a algumas questões de natureza *versus* criação são de importância prática. Se a pressão arterial alta ocorre primariamente em virtude de escolhas no estilo de vida (ambiente), então as alterações nos hábitos alimentares ou de exercícios são as mais apropriadas. Entretanto, se a pressão arterial alta for amplamente predeterminada por nossos genes, então a terapia medicamentosa pode ser recomendada.

Geneticistas quantitativos desenvolveram ferramentas estatísticas necessárias para estimar, com precisão razoável, a extensão na qual a variação em traços complexos ocorre em virtude dos genes *versus* do ambiente. A seguir, descreveremos essas ferramentas. No fim desta seção, discutiremos sobre as presunções subjacentes a essas estimativas e os limites da sua utilidade.

Iniciaremos definindo a **herdabilidade no sentido amplo** (H^2) como a parte da variância fenotípica que ocorre em virtude de diferenças genéticas entre os indivíduos em uma população. Matematicamente, escrevemos isso como a razão entre variância genética e a variância total na população:

$$H^2 = \frac{V_g}{V_X}$$

H está ao quadrado porque é a razão de duas variâncias, que são medidas em unidades ao quadrado. H^2 pode variar de 0 a 1,0. Quando toda a variação em uma população ocorre em virtude de fontes ambientais e não há variação genética, então H^2 é 0. Quando toda a variação em uma população ocorre em virtude de fontes genéticas, então V_g é igual a V_X e H^2 é 1,0. H^2 é denominada “no sentido amplo” porque envolve diversos diferentes modos por meio dos quais os genes contribuem para a variação. Por exemplo, uma parte da variação ocorrerá em virtude das contribuições de genes individuais. A contribuição da variação

genética adicional pode ocorrer em virtude do modo como os genes atuam em conjunto, das interações dos genes ou da epistasia.

Na Seção 19.2, demonstramos como podemos calcular as variâncias genética e ambiental quando temos linhagens endocruzadas ou clones. Em relação ao exemplo imaginário de dias até a polinização das linhagens endocruzadas de milho na Tabela 19.2 (experimento I), vimos que V_g é 12,0 e V_X é 14,67. Com a utilização desses valores, a herdabilidade do traço é 12,0/14,67 = 0,82 ou 82%. Essa estimativa de H^2 nos informa que os genes contribuem com a maior parte da variação e que os fatores ambientais contribuem com uma parte mais modesta da variação. Portanto, podemos concluir que os dias até a polinização são um traço altamente hereditário no milho.

Vejamos os dados em relação ao experimento II na Tabela 19.2. Os genótipos são exatamente os mesmos do experimento I; esses são os genótipos das linhagens endocruzadas A a J. Nesse caso, entretanto, as linhagens são criadas em ambientes mais extremos. Se calcularmos a variância em relação às médias da linhagem endocruzada no experimento II, V_g será 12,0 dias², como no experimento I. Tendo em vista que os genótipos são os mesmos em ambos os experimentos, a variância genética é a mesma. Se calcularmos a variância em relação às médias dos diferentes ambientes (V_e) no experimento II, obteremos 24,0 dias², que é muito maior do que o valor de V_e no experimento I (2,67). Tendo em vista que os ambientes são mais extremos, a variância ambiental é maior. Finalmente, se calcularmos H^2 em relação ao experimento II, obtemos:

Gêmeos monozigóticos são geneticamente idênticos

A estimativa de H^2 em relação ao experimento II é menor – mais próxima de 0 do que de 1. Portanto, podemos concluir que os dias até a polinização *não* são um traço altamente hereditário no milho.

O contraste entre as estimativas da herdabilidade em relação ao mesmo conjunto de linhagens endocruzadas de milho criadas em diferentes ambientes destaca o ponto de que a *herdabilidade é a proporção da variância fenotípica (V_X) devida à genética*. Tendo em vista que $V_X = V_g + V_e$, na medida em que V_e aumenta, V_g representará então uma parte menor de V_X e H^2 diminuirá. De modo semelhante, se a variância ambiental for mantida no mínimo, então V_g representará uma parte maior de V_X e H^2 aumentará. H^2 é um alvo em movimento e os resultados de um estudo podem não se aplicar a outro.

Medida da herdabilidade em seres humanos a partir de estudos com gêmeos

Como podemos medir a herdabilidade em seres humanos? Embora não tenhamos linhagens endocruzadas em seres humanos, temos indivíduos geneticamente idênticos – os gêmeos monozigóticos ou idênticos (Figura 19.5). Na maioria dos casos, os gêmeos idênticos são criados na mesma família e, assim, vivenciam um

ambiente semelhante. Quando indivíduos com os mesmos genótipos são criados nos mesmos ambientes, violamos a presunção de nosso modelo genético de que os genes e o ambiente são independentes. Assim, para estimar a herdabilidade em seres humanos, precisamos utilizar conjuntos de gêmeos idênticos que tenham sido separados logo após o nascimento e criados em separado por pais adotivos não relacionados.

A equação para a estimativa da H^2 em estudos de gêmeos idênticos que são criados em separado é relativamente simples. Ela utiliza a medida estatística denominada covariância, que foi introduzida no Quadro 19.2. Conforme explicado no Quadro 19.3, a covariância entre gêmeos idênticos que são criados em separado é igual à variância genética (V_g). Portanto, podemos estimar H^2 em seres humanos ao utilizar essa covariância como o numerador e a variância do traço (V_X) como o denominador:

$$H^2 = \frac{COV_{X',X''}}{V_X}$$

Aqui está como isso é feito. Para cada conjunto de gêmeos, designaremos o valor do traço em relação a um gêmeo como X' e ao outro, como X''. Se tivermos n conjuntos de gêmeos, então os valores dos traços em relação aos n conjuntos deverão ser designados X_1', X_1'', X_2', X_2''… X_n', X_n''.

FIGURA 19.5 Um par de gêmeos idênticos. (Barbara Penoyar/Photodisc/Getty Images.)

Suponha que tivéssemos as medições do QI em relação a cinco conjuntos de gêmeos, como segue:

	Gêmeo	
	X'	X''
1	100	110
2	125	118
3	97	90
4	92	104
5	86	89

Utilizando esses dados e a fórmula em relação à covariância do Quadro 19.2, calculamos que a $COV_{X',X''}$ é 119,2 pontos². Utilizando a fórmula em relação à variância do traço, calcularíamos que o valor de V_X é 154,3 pontos². Portanto, obtemos:

$$H^2 = \frac{119{,}2 \text{ pontos}^2}{154{,}3 \text{ pontos}^2} = 0{,}77$$

Os pontos² no numerador e no denominador são cancelados e ficamos com uma medida sem unidade, que é a proporção da variância total que ocorre em virtude da genética.

O Quadro 19.3 fornece alguns detalhes adicionais a respeito da estimativa de H^2 a partir dos dados de gêmeos, incluindo a derivação da fórmula que acabamos

de utilizar. Ele também discute a relação entre a razão de $COV_{X',X'}/V_X$ e o coeficiente de correlação. Geneticistas quantitativos desenvolveram diversos meios para estimar a herdabilidade com a utilização da correlação entre parentes. Gêmeos idênticos compartilham 100% de seus genes, enquanto irmãos, irmãs e gêmeos dizigóticos compartilham 50% de seus genes. A força da correlação entre os diferentes tipos de parentes pode ser dimensionada em relação à proporção dos genes que eles compartilham e os resultados, utilizados para estimar as contribuições genética e ambiental para a variação do traço.

Quadro 19.3 Estimativa da herdabilidade a partir de estudos com gêmeos humanos.

Se tivéssemos muitos conjuntos de gêmeos idênticos que fossem criados em separado, como poderíamos utilizá-los para medir H^2? Representaremos arbitrariamente o valor do traço em relação a um membro de cada par de gêmeos como X' e o valor do traço em relação ao outro como X''. Temos muitos (n) conjuntos de gêmeos: $X_1'X_1''$, $X_2'X_2''$... $X_n'X_n''$. Podemos expressar os desvios fenotípicos em relação a um grupo de gêmeos como a soma de seus desvios genético e ambiental.

$$x' = g + e' \text{ e } x'' = g + e''$$

utilizando x' como o desvio em relação a um gêmeo e x'' em relação ao outro gêmeo. Observe que g é o mesmo, tendo em vista que os gêmeos são geneticamente idênticos, mas e' e e'' são diferentes em virtude de os gêmeos terem sido criados em famílias separadas. Em seguida, desenvolvemos uma expressão para a covariância entre os gêmeos. No Quadro 19.2, vimos que a covariância é a média ou o valor esperado dos produtos cruzados $E(xy)$. Utilizando nossa notação em relação aos gêmeos, x' e x'', em vez de x e y, obtemos:

$$COV_{X',X''} = E(x'x'')$$

Podemos substituir ($g + e'$) por x' e ($g + e''$) por x'', que nos fornece:

$$\begin{aligned} COV_{X',X''} &= E[(g \times e')(g \times e'')] \\ &= E(g^2 + ge' + ge'' + e'e'') \\ &= E(g^2) + E(ge') + E(ge'') + E(e'e'') \end{aligned}$$

Consideraremos os últimos três termos dessa expressão. Sob nosso modelo, os gêmeos são atribuídos aleatoriamente a famílias e, portanto, não deve haver correlação entre os ambientes aos quais os gêmeos X' e X'' de cada par são atribuídos. De acordo com isso, a covariância entre os ambientes [$E(e'e'')$] será 0,0. De modo semelhante, tendo em vista que a atribuição dos gêmeos às famílias é aleatória, esperamos uma ausência de correlação entre o desvio genético dos gêmeos (g) e a família à qual eles são atribuídos, de modo que $E(ge')$ e $E(ge'')$ serão 0,0. Portanto, a equação em relação à covariância entre os gêmeos é reduzida para:

$$COV_{X',X''} = E(g^2) = V_g$$

Em outras palavras, a covariância entre gêmeos idênticos criados em separado é igual à variância genética. Se tivermos um grande conjunto de gêmeos idênticos que foram criados em separado, podemos utilizar a covariância entre os gêmeos para estimar a quantidade de variação genética em relação a um traço na população geral. Se dividirmos essa covariância pela variância fenotípica, teremos então uma estimativa de H^2:

$$H^2 = \frac{COV_{X',X''}}{V_X}$$

Essa equação é essencialmente o coeficiente de correlação entre os gêmeos. Espera-se que a variância em relação ao gêmeo de cada conjunto designado X' e aquela em relação ao gêmeo designado X'' sejam as mesmas em uma amostra grande. Portanto, podemos reescrever o denominador da equação como segue:

$$r_{X',X''} = \frac{COV_{X',X''}}{\sqrt{V_{X'}V_{X''}}} = H^2$$

e veremos que H^2 é equivalente à correlação entre gêmeos.

Ao longo dos últimos 100 anos, houve extensivos estudos genéticos com gêmeos e outros conjuntos de parentes. Foi aprendido muito a respeito da variação hereditária em seres humanos a partir desses estudos. A Tabela 19.4 lista alguns resultados de estudos com gêmeos. Pode ou não ser uma surpresa para você, mas existe uma contribuição genética para a variância em relação a muitos traços diferentes, incluindo a psique, a fisiologia, os atributos da personalidade, os transtornos psiquiátricos e até mesmo as nossas atitudes sociais e crenças políticas. Observamos prontamente que traços tais como a cor dos cabelos e dos olhos ocorrem nas famílias e sabemos que esses traços são a manifestação de processos bioquímicos e do desenvolvimento controlados geneticamente. Nesse contexto, não é tão surpreendente que outros aspectos de quem somos como pessoa também apresentem uma influência genética.

Tabela 19.4 Herdabilidade no sentido amplo em relação a alguns traços em seres humanos, conforme determinado por estudos com gêmeos.

Traço	H^2
Atributos físicos	
Altura	0,88
Circunferência torácica	0,61
Circunferência da cintura	0,25

Contagem de cristas digitais	0,97
Pressão arterial sistólica	0,64
Frequência cardíaca	0,49
Atributos mentais	
QI	0,69
Velocidade do processamento espacial	0,36
Velocidade da aquisição de informações	0,20
Velocidade do processamento de informações	0,56
Atributos da personalidade	
Extroversão	0,54
Conscientização	0,49
Neuroticismo	0,48
Emocionalidade positiva	0,50
Comportamento antissocial em adultos	0,41

Transtornos psiquiátricos	
Autismo	0,90
Esquizofrenia	0,80
Depressão maior	0,37
Transtornos de ansiedade	0,30
Alcoolismo	0,50 a 0,60
Crenças e atitudes políticas	
Religiosidade entre adultos	0,30 a 0,45
Conservadorismo entre adultos	0,45 a 0,65
Visões de orador escolar	0,41
Visões sobre pacifismo	0,38

Fontes: J. R. Alford *et al.*, *American Political Science Review* 99, 2005, 1-15; T. Bouchard *et al.*, *Science* 250, 1990, 223-228; T. Bouchard, *Curr. Dir. Phys. Sci.* 13, 2004, 148-151; P. J. Clark, *Am. J. Hum. Genet.* 7, 1956, 49-54; C. M. Freitag, *Mol. Psychiatry* 12, 2007, 2-22.

Estudos com gêmeos e as estimativas da herdabilidade que eles proporcionam

podem ser facilmente interpretados excessiva ou erroneamente. Aqui estão alguns pontos importantes para se ter em mente. Primeiramente, H^2 é uma propriedade de uma população e de um ambiente em particular. Por esse motivo, as estimativas de H^2 podem diferir amplamente entre populações e ambientes diferentes. Observamos esse fenômeno anteriormente, no caso dos dias até a polinização em relação às linhagens endocruzadas de milho. Em segundo lugar, os conjuntos de gêmeos utilizados em muitos estudos foram separados ao nascimento e inseridos em lares adotivos. As agências de adoção não atribuem os bebês aleatoriamente para uma ampla gama de famílias em uma sociedade; em vez disso, eles inserem os bebês em famílias econômica, social e emocionalmente estáveis. Como resultado, V_e é menor do que na população em geral e a estimativa de H^2 será inflada. De acordo com isso, as estimativas publicadas provavelmente nos levam a subestimar a importância do ambiente e a superestimar a importância da genética. Em terceiro lugar, em relação a gêmeos, os efeitos pré-natais podem causar uma correlação positiva entre o genótipo e o ambiente. Conforme vimos anteriormente no caso dos puros-sangues-ingleses e dos jóqueis, tal correlação viola o nosso modelo e desviará H^2 para cima.

Finalmente, a herdabilidade não é útil na interpretação das diferenças entre grupos. A Tabela 19.4 demonstra que a herdabilidade em relação à altura em seres humanos pode ser muito alta: 0,88. Entretanto, esse valor alto em relação à herdabilidade absolutamente não nos informa se grupos com alturas distintas diferem em virtude da genética ou do ambiente. Por exemplo, homens na Holanda atualmente têm em média 184 cm de altura, enquanto aproximadamente em 1800, os homens na Holanda tinham aproximadamente 168 cm de altura em média, uma diferença de 16 cm. O *pool* gênico dos holandeses provavelmente não foi alterado sensivelmente ao longo daquele período e, assim, a genética não pode explicar a enorme diferença na altura entre a população atual e aquela de 200 anos atrás. Em vez disso, melhoras na saúde e na nutrição provavelmente são a causa. Portanto, embora a altura seja altamente hereditária e as populações holandesas anteriores e atuais sejam muito diferentes na altura, a diferença apresenta uma base ambiental.

> CONCEITO-CHAVE A herdabilidade no sentido amplo (H^2) é a razão entre a variação genética (V_g) e a fenotípica (V_x). H^2 fornece uma medida do quanto as diferenças entre os indivíduos de uma população ocorrem em virtude de fatores genéticos *versus* ambientais. As estimativas de H^2 se aplicam apenas à população e ao ambiente no qual elas foram feitas. H^2 não é útil para interpretar diferenças nas médias dos traços entre populações.

19.4 Herdabilidade no sentido restrito | Previsão dos fenótipos

A herdabilidade no sentido amplo nos informa a proporção da variância em uma população que ocorre em virtude de fatores genéticos. A herdabilidade no sentido amplo expressa o grau até o qual as diferenças nos fenótipos entre os indivíduos em uma população são determinados por diferenças nos seus genótipos. Entretanto, até mesmo quando existe variação genética em uma população, conforme medida pela herdabilidade no sentido amplo, ela pode não ser transmissível para a próxima geração de modo previsível. Nesta seção, exploraremos como a variação genética se apresenta de dois modos – a variação aditiva e de dominância (não aditiva). Enquanto a variação aditiva é transmitida de modo previsível dos genitores para a descendência, a variação de dominância não é. Também definiremos outro tipo de herdabilidade, denominada **herdabilidade no sentido restrito**, que é a razão entre a variância aditiva e a variância fenotípica. A herdabilidade no sentido restrito fornece uma medida do grau em que a constituição genética dos indivíduos determina os fenótipos de sua descendência.

Os diferentes tipos de **ação gênica** (interação dos alelos em um *locus*) estão no cerne da compreensão da herdabilidade no sentido restrito e, assim, vamos revisá-los brevemente. Considere um *locus*, *B,* que controla o número de flores em uma planta. O *locus* apresenta dois alelos, B_1 e B_2, e três genótipos – B_1/B_1, B_1/B_2 e B_2/B_2. Conforme diagramado na Figura 19.6 A, plantas com o genótipo

B_1/B_1 apresentam 1 flor, plantas B_1/B_2 apresentam 2 flores e plantas B_2/B_2 apresentam 3 flores. Em um caso como esse, quando o valor do traço do heterozigoto é intermediário entre aqueles das duas classes homozigotas, a ação gênica é definida como **aditiva**. Na Figura 19.6 B, o heterozigoto apresenta 3 flores, o mesmo que o homozigoto B_2/B_2. Aqui, o alelo B_2 é dominante em relação ao alelo B_1. Nesse caso, a ação gênica é definida como **dominante**. (Também poderíamos definir essa ação gênica como **recessiva**, com o alelo B_1 sendo recessivo em relação ao alelo B_2.) A ação gênica não precisa ser puramente aditiva ou dominante, mas pode demonstrar **dominância parcial**. Por exemplo, se heterozigotos B_1/B_2 apresentassem 2,5 flores em média, diríamos então que o alelo B_2 demonstra dominância parcial.

Ação gênica e transmissão de variação genética

Trabalharemos com um exemplo simples para demonstrar como o modo de ação gênica influencia a herdabilidade. Suponha que um floricultor deseje criar uma população de plantas melhoradas, com mais flores por planta. O número de flores é controlado pelo *locus B*, que apresenta dois alelos, B_1 e B_2, conforme diagramado na Figura 19.6 A. As frequências dos alelos B_1 e B_2 são, ambas, 0,5 e as frequências dos genótipos B_1/B_1, B_1/B_2 e B_2/B_2 são de 0,25, 0,50 e 0,25, respectivamente. Plantas com o genótipo B_1/B_1 apresentam 1 flor, plantas B_1/B_2 apresentam 2 flores e plantas B_2/B_2 apresentam 3 flores. O número médio de flores por planta na população é 2,0. (Lembre que podemos calcular a média como a soma dos produtos da frequência de cada classe vezes o valor em relação àquela classe.)

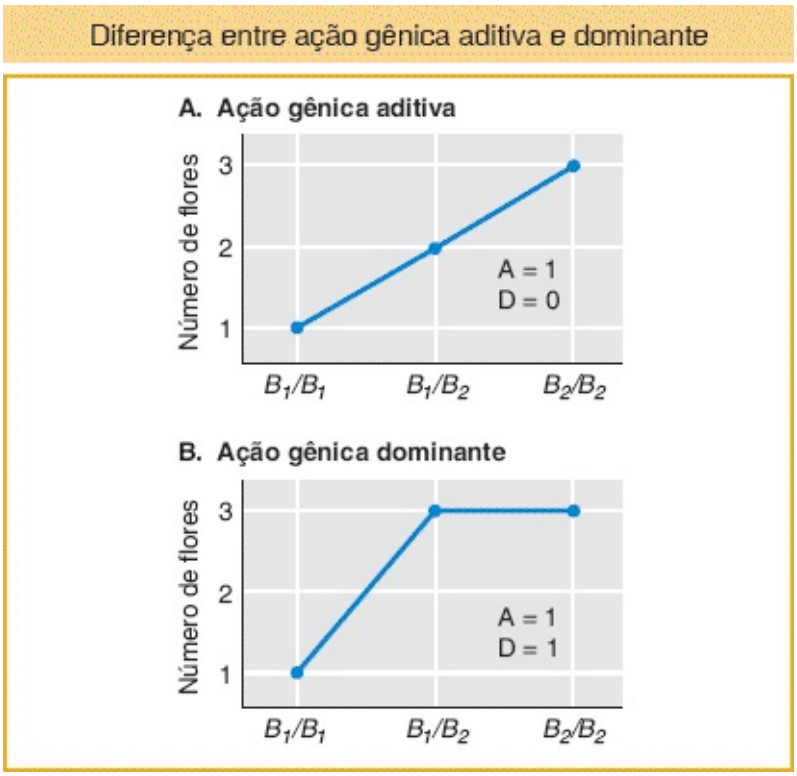


FIGURA 19.6 Gráfico do genótipo (eixo *x*) pelo fenótipo (eixo *y*) em relação a um *locus* hipotético, *B*, que regula o número de flores por planta. **A.** Ação gênica aditiva. **B.** Ação gênica dominante.

Genótipo	Frequência	Valor do traço (nº de flores)	Contribuição para a média (frequência × valor)
B_1/B_1	0,25	1	0,25
B_1/B_2	0,50	2	1,0
B_2/B_2	0,25	3	0,75
			Média = 2,0

Tendo em vista que o heterozigoto apresenta um fenótipo intermediário entre as duas classes homozigotas, a ação gênica é aditiva. Não existem efeitos ambientais e o genótipo isoladamente determina o número de flores, de modo que H^2 é 1,0. Se o cultivador de plantas selecionar plantas com 3 flores (B_2/B_2), realizar o seu intercruzamento e em seguida cultivar a descendência, então todas a descendência será B_2/B_2 e o número médio de flores por planta na descendência será 3,0. Quando a ação gênica é completamente aditiva e não existem efeitos ambientais, o fenótipo é totalmente hereditário. A seleção, conforme praticada pelo cultivador de plantas, atua perfeitamente.

Agora consideraremos o caso diagramado na Figura 19.6 B, na qual o alelo B_2 é dominante em relação ao B_1. Nesse caso, o heterozigoto B_1B_2 tem 3 flores. A frequência de ambos os alelos B_1 e B_2 é de 0,5 e as frequências dos genótipos B_1/B_1, B_1/B_2 e B_2/B_2 são de 0,25, 0,50 e 0,25, respectivamente. Novamente, não existe contribuição ambiental para as diferenças entre os indivíduos, de modo que H^2 é 1,0. O número médio de flores por planta na população inicial é 2,5.

Genótipo	Frequência	Fenótipo	Contribuição para a média (frequência × valor)
B_1/B_1	0,25	1	0,25
B_1/B_2	0,50	3	1,5
B_2/B_2	0,25	3	0,75
			Média = 2,5

Se o floricultor selecionar um grupo de plantas com 3 flores, 2/3 serão B_1/B_2 e

1/3, B_2/B_2. Quando o floricultor realizar o intercruzamento das plantas selecionadas, 0,44 (2/3 × 2/3) dos cruzamentos serão entre heterozigotos e 1/4 da descendência desses cruzamentos será B_1/B_1 e, portanto, com 1 flor. O remanescente da descendência será B_1/B_2 ou B_2/B_2 e, portanto, com 3 flores. A média geral em relação à descendência será de 2,78, embora a média de seus genitores tenha sido de 3,0. Portanto, quando existe dominância, o fenótipo não é totalmente hereditário. A seleção, conforme praticada pelo floricultor, atuou, porém não perfeitamente, tendo em vista que algumas das diferenças entre os indivíduos ocorrem em virtude da dominância.

Concluindo, quando existe dominância, não podemos prever estritamente os fenótipos da descendência a partir dos fenótipos dos genitores. Algumas das diferenças (variação) entre os indivíduos na geração parental ocorrem em virtude de interações de dominância entre os alelos. Tendo em vista que os genitores transmitem os seus genes, mas não os seus genótipos para a sua descendência, essas interações de dominância não são transmitidas para a descendência.

Efeitos aditivos e da dominância

Conforme descrito anteriormente, os traços controlados por genes com ação gênica aditiva responderão de modo muito diferente daqueles com dominância. Portanto, os geneticistas precisam quantificar o grau de dominância e aditividade. Nesta seção, veremos como isso é feito. Consideraremos novamente o *locus B* que controla o número de flores em uma planta (ver Figura 19.6). O **efeito aditivo (A)** proporciona uma medida do grau de alteração no fenótipo que ocorre com a substituição de um alelo B_2 por um alelo B_1. O efeito aditivo é calculado como a diferença entre as duas classes homozigotas dividida por 2. Por exemplo, conforme demonstrado na Figura 19.6 A, se o valor do traço do genótipo B_1/B_1 for 1 e o valor do traço do genótipo B_2/B_2 for 3, então:

$$\mathrm{A} = \frac{X_{B_2B_2} - X_{B_1B_1}}{2} = \frac{3-1}{2} = 1$$

O **efeito da dominância (D)** é o desvio do heterozigoto (B_1/B_2) do ponto médio

das duas classes homozigotas. Conforme demonstrado na Figura 19.6 B, se o valor do traço do genótipo B_1/B_1 for 1, do genótipo B_2/B_2, 3, e do genótipo B_2/B_2, 3, então:

Se você calcular D em relação à situação ilustrada na Figura 19.6 A, você encontrará D = 0; ou seja, nenhuma dominância.

$$D = X_{B_1B_2} - \left(\frac{X_{B_2B_2} + X_{B_1B_1}}{2} \right) = 3 - 2 = 1$$

A razão de D/A fornece uma medida do grau de dominância. Para a Figura 19.6 A, D/A = 0,0, indicando aditividade pura ou nenhuma dominância. Para a Figura 19.6 B, D/A = 1,0, indicando dominância completa. Uma razão de D/A de –1 indicaria um recessivo completo. (A distinção entre dominância e recessividade depende do modo como os fenótipos são codificados e, nesse sentido, é arbitrária.) Os valores que são superiores a 0 e inferiores a 1 representam dominância parcial e os valores que são inferiores a 0 e superiores a –1 representam recessividade parcial.

Aqui está um exemplo de cálculo dos efeitos aditivos e da dominância em um único *locus*. O peixe esgana-gato (*Gasterosteus aculeatus*) apresenta populações marinhas com longas espinhas pélvicas e populações que vivem próximas do fundo de lagos de água doce, com espinhas pélvicas altamente reduzidas (Figura 19.7 A). Acredita-se que as espinhas desempenhem um papel na defesa contra a predação. As populações que vivem no fundo de água doce são derivadas de populações marinhas ancestrais. Uma alteração na predação entre os ambientes marinho e de água doce pode explicar a perda das espinhas nos ambientes de água doce (ver Capítulo 20).

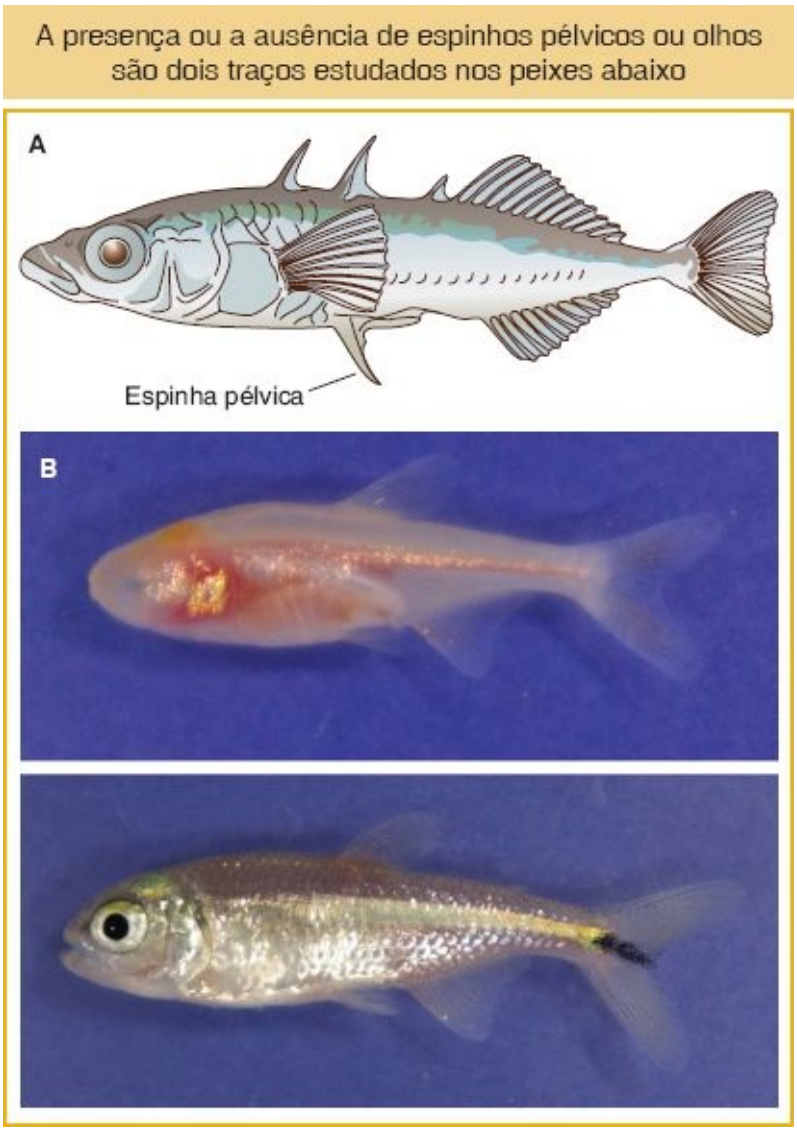


FIGURA 19.7 A. O esgana-gato (*Gasterosteus aculeatus*) com três espinhas dorsais. **B.** O peixe-cego (*Astyanax mexicanus*) de cavernas (*parte superior*) e seu parente da superfície com visão (*parte inferior*). (*B.* Masato Yoshizawa e William Jeffery, University of Maryland.)

O *Pitx1* é um de diversos genes que contribuem para o comprimento das espinhas pélvicas nos esgana-gatos. Esse gene codifica um fator de transcrição que regula o desenvolvimento da pelve em vertebrados, incluindo o crescimento de espinhas pélvicas em esgana-gatos. Michael Shapiro e seus colegas na Stanford University mediram o comprimento da espinha pélvica em uma população F_2 que segregava o alelo marinho ou longo (*l*) e o alelo de água doce ou curto (*s*) do *Pitx1*. Eles registraram os valores médios a seguir (em unidades

proporcionais ao comprimento corporal) em relação ao comprimento da espinha pélvica para as três classes genotípicas:

s/s	*s/l*	*l/l*
0,068	0,132	0,148

Utilizando esses valores e as fórmulas anteriores, podemos calcular os efeitos aditivos e da dominância. O efeito aditivo (A) é:

$$(0{,}148 - 0{,}068)/2 = 0{,}04$$

ou 4% do comprimento corporal. O efeito da dominância (D) é:

$$0{,}132 - [(0{,}148 + 0{,}068)/2] = 0{,}024$$

A razão de dominância/aditividade é:

$$0{,}024/0{,}04 = 0{,}6$$

O valor de 0,6 em relação à razão indica que o alelo longo (*l*) de *Pitx1* é parcialmente dominante em relação ao alelo curto (*s*).

Pode-se também calcular a média dos efeitos aditivos e da dominância sobre todos os genes no genoma que afetam o traço. Aqui está um exemplo com a utilização do peixe-cego (*Astyanax mexicanus*) de cavernas e seus parentes da superfície (Figura 19.7 B). As populações de cavernas apresentam olhos altamente reduzidos (de pequeno diâmetro) em comparação às populações da superfície. As populações que colonizam cavernas sem luz não se beneficiam do fato de ter olhos. Tendo em vista que existem custos fisiológicos e neurológicos para a formação e a manutenção dos olhos, a evolução pode ter forçado uma redução no tamanho dos olhos nas populações de caverna.

Horst Wilkins na University of Hamburg mediu o diâmetro médio dos olhos

(em mm) em populações de caverna e da superfície e de seu híbrido F_1:

Caverna	F_1	Superfície
2,10	5,09	7,05

Com a utilização das fórmulas anteriores, calculamos que A = 2,48, D = 0,52 e D/A = 0,21. Nesse caso, a ação gênica está mais próxima de um estado puramente aditivo, embora o genoma dos peixes de superfície seja ligeiramente dominante.

> CONCEITO-CHAVE Quando o valor do traço em relação à classe heterozigota é intermediário entre as duas classes homozigotas, a ação gênica é denominada aditiva. Qualquer desvio do heterozigoto do ponto médio entre as duas classes homozigotas indica um grau de dominância de um alelo. Os efeitos aditivos (A), da dominância (D) e a sua razão (D/A) proporcionam medidas para quantificar o modo de ação gênica.

Modelo com aditividade e dominância

O exemplo anterior com o *locus B* e o número de flores demonstra que não podemos prever com precisão os fenótipos da descendência a partir dos fenótipos parentais quando existe dominância, embora possamos fazer isso em casos de aditividade pura. Ao prever os fenótipos da descendência, precisamos separar as contribuições aditiva e da dominância. Para tanto, precisamos modificar o modelo simples introduzido na Seção 19.2, $x = g + e$.

Iniciaremos ao observar mais de perto a situação ilustrada na Figura 19.6 B. Indivíduos com os genótipos B_1/B_2 e B_2/B_2 apresentam o mesmo fenótipo, 3 flores. Se subtrairmos a média da população (2,5) de seu valor de traço (3), observaremos que eles apresentam o mesmo desvio genotípico (g):

$$g_{B1B2} = g_{B2B2} = 0{,}5$$

Agora calcularemos os fenótipos médios de sua descendência. Se autopolinizarmos um indivíduo B_1/B_2, a descendência será $\frac{1}{4}$ B_1/B_1, $\frac{1}{2}$ B_1/B_2 e $\frac{1}{4}$ B_2/B_2 e o valor médio do traço dessa descendência será 2,75. Entretanto, se autopolinizarmos um indivíduo B_2/B_2, toda a descendência será B_2/B_2 e o valor médio do traço dessa descendência será 3,0. Embora os indivíduos B_1/B_2 e B_2/B_2 apresentem o mesmo valor de traço e o mesmo valor em relação ao seu desvio genotípico (g), eles não produzem descendência equivalente, tendo em vista que o fundamento de base de seus fenótipos é diferente. O fenótipo do indivíduo B_1/B_2 depende do efeito da dominância (D), enquanto aquele do indivíduo B_2/B_2 não envolve dominância.

Podemos expandir o modelo simples ($x = g + e$) para incorporar as contribuições aditiva e da dominância. O desvio genotípico (g) é a soma de dois componentes: a, o desvio aditivo, que é transmitido para a descendência, e d, o desvio de dominância, que não é transmitido para a descendência. Podemos reescrever o modelo simples e separar esses dois componentes como segue:

$$x = g + e$$
$$\downarrow$$
$$x = a + d + e$$

O desvio aditivo é transmitido dos genitores para a descendência de modo previsível. O desvio da dominância não é transmitido dos genitores para a descendência, tendo em vista que são criados novos genótipos e, portanto, novas interações dos alelos a cada geração.

Vejamos como o desvio genético é decomposto nos desvios aditivo e da dominância em relação ao caso demonstrado na Figura 19.6 B.

	B_1B_1	B_2B_2	B_2B_2
Valor de traço	1	3	3

Desvio genético (*g*)	–1,5	0,5	0,5
Desvio aditivo (*a*)	–1	0	1
Desvio da dominância (*d*)	–0,5	0,5	–0,5

Os desvios genotípicos (*g*) são calculados simplesmente subtraindo-se a média da população (2,5) do valor de traço em relação a cada genótipo. Cada desvio genotípico em seguida é decomposto nos desvios aditivo (*a*) e da dominância (*d*) com a utilização de fórmulas que estão além do escopo deste livro. Essas fórmulas incluem os efeitos aditivos (A) e de dominância (D), bem como as frequências do alelo B_1 e B_2 na população. Você observará que $a + d$ somam g. Os desvios aditivo (*a*) e da dominância (*d*) são dependentes das frequências dos alelos, tendo em vista que o fenótipo de uma descendência que recebe um alelo B_1 de um genitor dependerá da combinação desse alelo com um alelo B_1 ou B_2 do outro genitor e esse desfecho depende das frequências dos alelos na população.

O desvio aditivo (*a*) tem um significado importante no cultivo de plantas e na criação de animais. Ele é o **valor produtivo**, ou a parte do desvio de um indivíduo da média da população que ocorre em virtude de efeitos aditivos. Essa é a parte transmitida para sua progênie. Portanto, se desejarmos aumentar o número de flores por planta na população, os indivíduos B_2/B_2 apresentam o mais alto valor produtivo. Os valores produtivos também podem ser calculados em relação ao genoma em geral para um indivíduo. Criadores de animais estimam os valores produtivos genômicos de cada animal e essas estimativas podem determinar o valor econômico do animal.

Separamos o desvio genético (*g*) em desvios aditivo (*a*) e da dominância (*d*). Utilizando a álgebra semelhante àquela descrita no Quadro 19.2, também podemos separar a variância genética em variâncias aditiva e da dominância, como segue:

$$V_g = V_a + V_d$$

em que V_a é a **variância aditiva** e V_d é a variância da dominância. V_a é a variância dos desvios aditivos ou a variância dos valores genéticos. É a parte da variação genética que é transmitida dos genitores para a sua descendência. V_d é a variância dos desvios da dominância. Finalmente, podemos substituir esses termos na equação em relação à variância fenotípica apresentada anteriormente no capítulo:

$$V_X = V_g + V_e$$
$$V_X = V_a + V_d + V_e$$

em que V_e é a variância ambiental. Essa equação presume que os componentes aditivo e da dominância não estão correlacionados com os efeitos ambientais. Essa presunção será verdadeira em experimentos nos quais os indivíduos são atribuídos aleatoriamente aos ambientes.

Até agora, descrevemos modelos com variâncias e desvios genéticos, ambientais, aditivos e da dominância. Na genética quantitativa, os modelos podem chegar a ser até mesmo mais complexos. Em particular, os modelos podem ser expandidos para incluir a interação dos fatores. Se um fator altera o efeito de outro fator, existe então uma interação. O Quadro 19.4 revisa brevemente como as interações são levadas em conta nos modelos genéticos quantitativos.

CONCEITO-CHAVE O desvio genético (g) de um indivíduo da média da população é composto por duas partes – seu desvio aditivo (a) e seu desvio da dominância (d). O desvio aditivo é conhecido como valor genético e representa o componente do fenótipo de um indivíduo que é transmitido para a sua descendência.

A variação genética em relação a um traço em uma população (V_g) pode ser decomposta nas variâncias aditiva (V_a) e da dominância (V_d). A variância aditiva é a fração da variação genética que é transmitida dos genitores para a descendência.

Herdabilidade no sentido restrito

Agora podemos definir a *herdabilidade no sentido restrito*, que é simbolizada por uma letra minúscula *h* ao quadrado (h^2), como a razão entre a variância aditiva a variância fenotípica total:

$$h^2 = \frac{V_a}{V_X} = \frac{V_a}{V_a + V_d + V_e}$$

Esse tipo de herdabilidade mede o quanto da variação entre os indivíduos em uma população é transmitido de modo previsível para a sua descendência. A herdabilidade no sentido restrito é o tipo de herdabilidade de interesse para cultivadores de plantas e criadores de animais, tendo em vista que proporciona uma medida de quão bem um traço responderá ao cultivo e à criação seletiva.

Para estimar h^2, precisamos medir V_a, mas como isso pode ser realizado? Com a utilização da álgebra e da lógica semelhantes àquelas que utilizamos para demonstrar que V_g pode ser estimada com a utilização da covariância entre gêmeos monozigóticos criados em separado (ver Quadro 19.3), também pode ser demonstrado que a covariância entre um genitor e a sua descendência é igual à metade da variância aditiva:

$$COV_{P,O} = \frac{1}{2}V_a$$

A covariância dos genitores para a descendência é metade de V_a em virtude de a descendência herdar apenas metade de seus genes de cada genitor. Combinando essa fórmula com aquela em relação a h^2, obtemos:

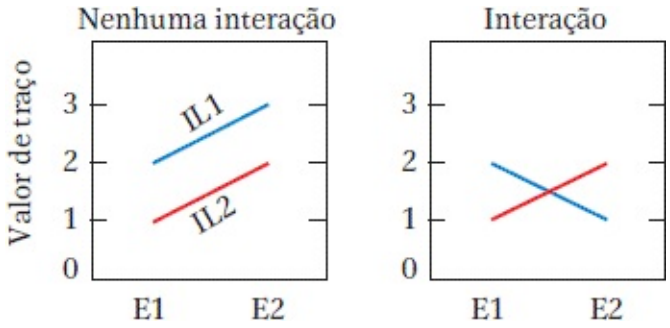
$$h^2 = \frac{V_a}{V_X} = \frac{2COV_{P,O}}{V_X}$$

Quadro 19.4 Efeitos da interação.

O modelo simples para a decomposição de traços em desvios genéticos e ambientais, $x = g + e$, presume que não existe interação do genótipo com o

ambiente. Por meio dessa declaração, queremos dizer que as diferenças entre os genótipos não são alteradas nos ambientes. Em outras palavras, uma *interação do genótipo com o ambiente* ocorre quando o desempenho de diferentes genótipos é afetado de modo desigual por uma alteração no ambiente. Aqui está um exemplo. Considere duas linhagens endocruzadas, IL1 e IL2, que apresentam genótipos diferentes. Criamos ambas as linhagens endocruzadas em dois ambientes, E1 ou E2. Podemos visualizar o desempenho dessas duas linhagens nos dois ambientes utilizando um gráfico (a seguir). Esse tipo de gráfico, que demonstra o padrão dos valores de traço de diferentes genótipos entre dois ou mais ambientes, é denominado *norma de reação*.

Se não houver interação, a diferença no valor de traço entre as linhagens endocruzadas será então a mesma em ambos os ambientes, conforme demonstrado pelo gráfico à esquerda.



Sem interação, a diferença entre as duas linhagens endocruzadas é 1,0 em ambos os ambientes e, assim, a diferença entre a média das linhagens nos dois ambientes é 1,0.

Ambiente 1: IL1 – IL2 = 2 – 1 = 1,0
Ambiente 2: IL1 – IL2 = 3 – 2 = 1,0

A diferença na média geral demonstra que as linhagens são geneticamente diferentes. A média em ambos os ambientes é de 2,5 para IL1 e 1,5 para IL2.

O gráfico à direita demonstra um caso de uma interação do genótipo com o ambiente. IL1 apresenta bom desempenho no Ambiente 1, mas desempenho

inadequado no Ambiente 2. O oposto é verdadeiro em relação a IL2. A diferença no valor de traço entre as duas linhagens é de +1,0 no Ambiente 1, mas –1,0 no Ambiente 2.

Ambiente 1: IL1 – IL2 = 2 – 1 = +1,0
Ambiente 2: IL1 – IL2 = 1 – 2 = –1,0

A diferença entre a média das linhagens nos dois ambientes é de 0,0 e, assim, podemos concluir erroneamente que essas linhagens endocruzadas são geneticamente equivalentes se observamos apenas a média geral.

O modelo simples pode ser expandido para incluir um termo da interação do genótipo com o ambiente ($g \times e$):

$$x = g + e + g \times e$$

e:

$$V_X = V_g + V_e + V_{g \times e}$$

em que $V_{g \times e}$ é a variância da interação do genótipo com o ambiente. Se o termo interação não estiver incluído no modelo, existe então uma presunção implícita de que não existem interações do genótipo com o ambiente.

As interações também podem ocorrer entre os alelos em genes separados. Esse tipo de interação é denominado epistasia. Vejamos como as interações epistáticas afetam a variação nos traços quantitativos.

Considere dois genes, *A* com os alelos A_1 e A_2 e *B* com os alelos B_1 e B_2. O lado esquerdo da tabela a seguir demonstra o caso de nenhuma interação desses genes. Iniciando com o genótipo A_1/A_1; B_1/B_1, sempre que você substituir um alelo A_2 por um alelo A_1, o valor do traço subirá em 1, independentemente do genótipo no *locus B*. O mesmo é verdadeiro quando substituímos os alelos no *locus B*. Os efeitos dos alelos no *locus A* são independentes daqueles no *locus B* e vice-versa. Não existe interação ou epistasia.

Nenhuma interação				Interação			
	B_1/B_1	B_1/B_2	B_2/B_2		B_1/B_1	B_1/B_2	B_2/B_2
A_1/A_1	0	1	2	A_1/A_1	0	1	2
A_1/A_2	1	2	3	A_1/A_2	0	1	3
A_2/A_2	2	3	4	A_2/A_2	0	1	4

Observe agora o lado direito da tabela. Iniciando com o genótipo A_1/A_1; B_1/B_1, a substituição de um alelo A_2 por um alelo A_1 apresenta um efeito apenas sobre o valor de traço quando o genótipo no *locus B* é B_2/B_2. Os efeitos dos alelos no *locus A* são *dependentes* daqueles no *locus B*. Existe uma interação ou epistasia entre os genes.

O modelo genético pode ser expandido para incluir um termo epistático, ou de interação (i):

$$x = a + d + i + e$$

e:

$$V_X = V_a + V_d + V_i + V_e$$

em que V_i é a variância da interação, ou variância epistática.

Se o termo interação não estiver incluído no modelo, existe então uma presunção implícita de que os genes atuam independentemente; ou seja, não existe epistasia. A variância da interação (V_i), assim como a variância da dominância, não é transmitida dos genitores para a sua descendência, tendo em vista que novos genótipos e, portanto, novas relações epistáticas são formados a cada geração.

Para estimar a V_a com a utilização da covariância entre os genitores e a descendência, é necessário controlar os fatores ambientais nos experimentos. Isso pode ser um desafio, tendo em vista que os genitores e a descendência

necessariamente são criados em ocasiões diferentes. V_a também pode ser estimada com a utilização da covariância entre meios-irmãos, caso em que todos os indivíduos no experimento podem ser criados ao mesmo tempo no mesmo ambiente. Meios-irmãos compartilham um quarto de seus genes e, assim, V_a é igual a 4 × a covariância entre meios-irmãos.

Se você comparar a equação de h^2 à equação de H^2 (ver Quadro 19.3), você observará que ambas envolvem a razão de uma covariância e uma variância. O coeficiente de correlação introduzido anteriormente no capítulo também é a razão de uma covariância e uma variância. Estamos utilizando o grau de correlação entre parentes para inferir a extensão até a qual os traços são hereditários.

Aqui está um exercício que a sua classe pode tentar fazer. Faça com que cada estudante apresente a sua altura e a altura do seu genitor do mesmo sexo. Com a utilização desses dados e de um *software* de planilhas, calcule a covariância entre os genitores e a sua descendência (os estudantes). Em seguida, estime h^2 como duas vezes a covariância dividida pela variância fenotípica. Para a variância fenotípica total (V_X) no denominador da equação, você pode utilizar a variância entre os genitores. Os dados em relação aos estudantes do sexo masculino e feminino devem ser analisados em separado.

Tipicamente, os valores em relação à herdabilidade no sentido restrito da altura em seres humanos são de aproximadamente 0,8, o que significa que aproximadamente 80% da variância é aditiva ou transmissível do genitor para a descendência. Os resultados da sua classe podem se desviar desse valor por diversos motivos. Primeiramente, se a sua classe for pequena, o erro de amostragem pode afetar a precisão da sua estimativa de h^2. Em segundo lugar, você não estará conduzindo um experimento randomizado. Se os genitores recriarem em suas famílias os ambientes de promoção do crescimento (ou de limitação do crescimento) que vivenciaram quando crianças, haverá então uma correlação entre os ambientes dos genitores e os de sua descendência. Essa correlação de ambientes viola uma presunção da análise. Em terceiro lugar, a população de estudantes em sua classe pode não ser representativa da população na qual o valor de 0,8 é obtido.

A Figura 19.8 é um gráfico de dispersão com os dados da altura de estudantes

dos sexos masculino e feminino e seus genitores. Existe uma clara correlação entre as alturas dos estudantes e de seu genitor do mesmo sexo. Esses dados fornecem estimativas da herdabilidade no sentido restrito de 0,86 para mãe e filha e de 0,82 para pai e filho. Os resultados estão próximos do valor de h^2 igual a 0,8 obtido a partir de estudos nos quais as crianças foram separadas de seus pais ao nascimento e criadas em famílias adotivas.

Aqui estão alguns pontos a mais a respeito da herdabilidade no sentido restrito. Primeiramente, quando $h^2 = 1{,}0$ ($V_a = V_X$), o valor esperado em relação ao fenótipo de uma descendência será igual ao valor de um dos genitores. Toda a variação na população é aditiva e herdada no sentido restrito. Em segundo lugar, quando $h^2 = 0{,}0$ ($V_a = 0$), o valor esperado do fenótipo de qualquer descendência será a média da população. Toda a variação na população ocorre em virtude de dominância ou de fatores ambientais e, portanto, não é transmissível para a descendência. Finalmente, assim como a herdabilidade no sentido amplo (H^2), a herdabilidade no sentido restrito é a propriedade do ambiente específico e da população em que foi medida. Uma estimativa de uma população e um ambiente pode não ser significativa para outra população ou outro ambiente.

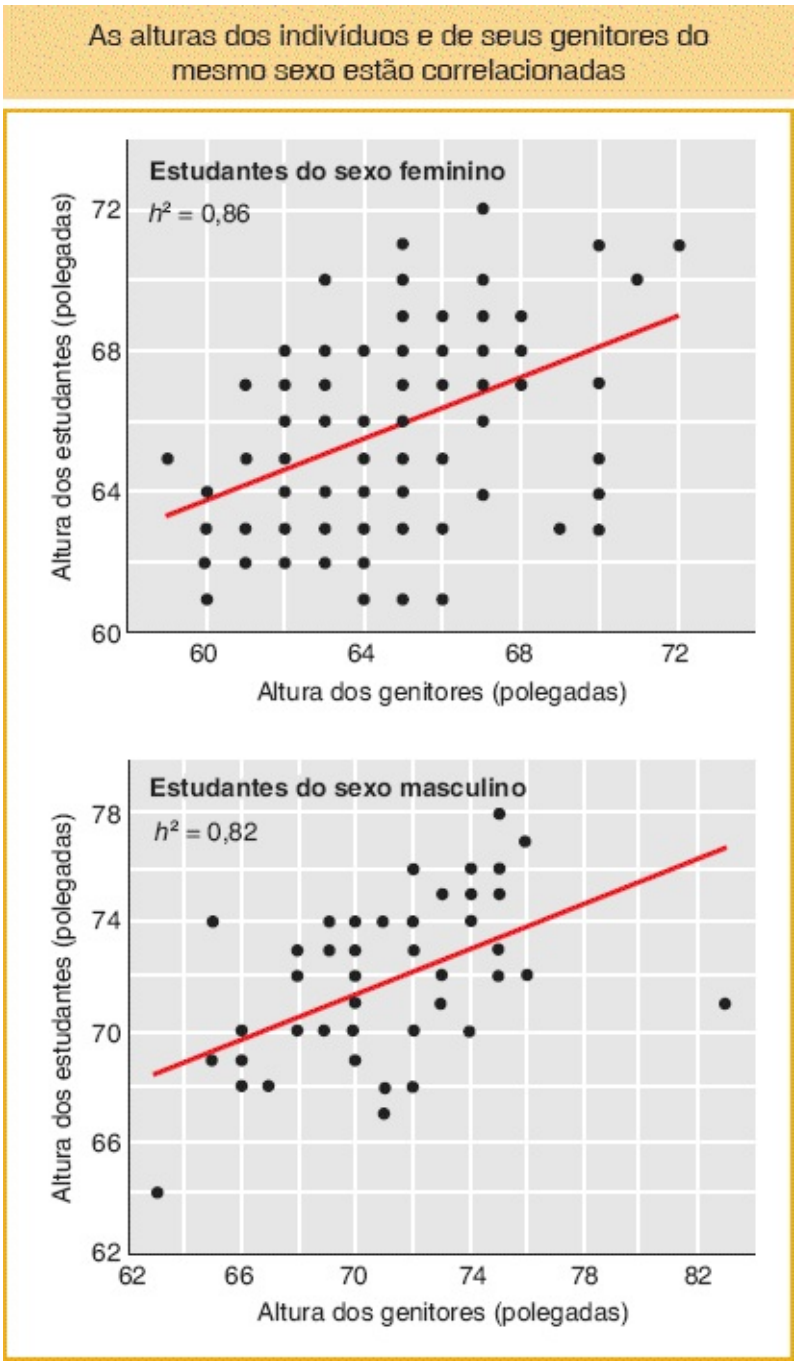


FIGURA 19.8 Diagramas de dispersão em relação à altura, em polegadas, de estudantes do sexo feminino (*parte superior*) e do sexo masculino (*parte inferior*) e seus genitores do mesmo sexo. Os gráficos demonstram correlações positivas entre as alturas dos estudantes e de seus genitores. A inclinação da linha diagonal é igual ao coeficiente de correlação.

A herdabilidade no sentido restrito é um conceito importante no cultivo de plantas e na criação de animais e na evolução. Para um cultivador ou criador, h^2

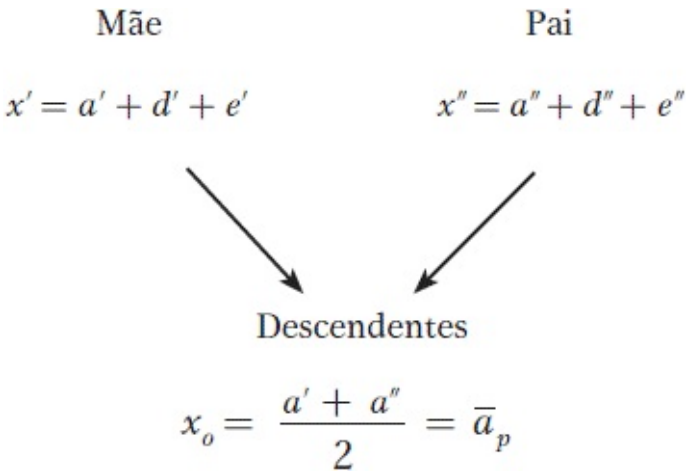
indica quais traços podem ser melhorados por meio de seleção artificial. Para um biólogo evolutivo, h^2 é crítica para compreender como as populações serão alteradas em resposta à seleção natural imposta por um ambiente em alteração. A Tabela 19.5 lista as estimativas da herdabilidade no sentido restrito em relação a alguns traços e organismos.

Previsão dos fenótipos da descendência

Com a finalidade de melhorar os cultivos e as criações animais em relação aos traços de importância agronômica, o cultivador ou criador deve ser capaz de prever o fenótipo da descendência a partir dos fenótipos de seus genitores. Tais previsões são feitas com a utilização do conhecimento do cultivador ou criador sobre a herdabilidade no sentido restrito. O desvio fenotípico de um indivíduo (x) da média da população é a soma dos desvios aditivo, de dominância e ambiental:

$$x = a + d + e$$

A parte aditiva é a parte herdada que é transmitida para a descendência. Veremos um conjunto de genitores com desvios fenotípicos x' em relação à mãe e x'' em relação ao pai. Os desvios de dominância dos genitores (d' e d'') não são transmitidos para a sua descendência, tendo em vista que são criados novos genótipos e novas interações de dominância a cada geração. De modo semelhante, os genitores não transmitem seus desvios ambientais (e' e e'') para a sua descendência.



Portanto, os únicos fatores que os genitores transmitem para a sua descendência são os seus desvios aditivos (a' e a''). De acordo com isso, podemos estimar o desvio fenotípico da descendência (x_0) como a média dos desvios aditivos de seus genitores (a_p).

Assim, para prever o fenótipo da descendência, precisamos conhecer os desvios aditivos dos genitores. Não podemos observar diretamente os desvios aditivos dos genitores, mas podemos estimá-los. O desvio aditivo de um indivíduo é a parte herdada de seu desvio fenotípico; ou seja:

$$\hat{a} = h^2 x$$

em que $\hat{a}$ significa uma estimativa do desvio aditivo ou do valor genético. Portanto, podemos estimar a média dos desvios aditivos dos genitores como o produto de h^2 vezes a média de seu desvio fenotípico e esse produto será uma estimativa do desvio fenotípico da descendência ($\hat{x}_0$):

Tabela 19.5 Herdabilidade no sentido restrito em relação a alguns traços em diversas espécies diferentes.

Traço	***h*² (%)**
Espécies agronômicas	
Peso corporal em bovinos	65
Produção leiteira em bovinos	35
Espessura da capa de gordura em suínos	70
Tamanho da ninhada em suínos	5

Peso corporal em galinhas	55
Peso do ovo em galinhas	50
Espécies naturais	
Comprimento do bico em tentilhões-de-darwin	65
Duração do voo no *Oncopeltus fasciatus*	20
Altura das plantas *Impatiens biflora* e *I. pallida*	8
Fecundidade do *Cervus elaphus*	46
Período de vida do pássaro *Ficedula albicollis*	15

Fonte: D. F. Falconer e T. F. C. Mackay, *Introduction to Quantitative Genetics*, Longman, 1996; J. C. Conner e D. L. Hartl, *A Primer in Ecological Genetics*, Sinauer, 2004.

$$\hat{x}_o = h^2\left(\frac{x' + x''}{2}\right)$$

ou:

$$\hat{x}_0 = h^2\overline{x}_p$$

A descendência apresentará seus próprios desvios de dominância e ambiental. Entretanto, esses não podem ser previstos. Tendo em vista que são desvios, em média, eles serão zero sobre um grande número de descendentes.

Aqui está um exemplo. Ovelhas islandesas são premiadas pela qualidade de

seu velo. A ovelha adulta média em uma população particular produz 2,7 kg de velo por ano. Um carneiro que produz 2,9 kg por ano é cruzado com uma ovelha que produz 3,2 kg por ano. A herdabilidade no sentido restrito da produção de velo nessa população é 0,4. Qual é a produção de velo prevista para a descendência desse cruzamento? Primeiramente, calcule os desvios fenotípicos em relação aos genitores subtraindo a média da população de seus valores fenotípicos:

Carneiro	6,5 – 6,0 = 0,5
Ovelha	7,0 – 6,0 = 1,0
Média dos genitores ($-x_p$)	(0,5 + 1,0)/2 = 0,75

Agora multiplique h^2 vezes $-x_p$ para determinar $\hat{x}_0$, o desvio fenotípico estimado da descendência:

$$0,4 \times 0,75 = 0,3$$

Finalmente, adicione a média da população (6,0) ao desvio fenotípico previsto da descendência (0,3) e obtenha o resultado de que o fenótipo previsto da descendência é 2,8 kg de velo por ano.

Pode parecer surpreendente a previsão de que a descendência produzirá menos velo do que os seus genitores. Entretanto, esse desfecho é esperado em relação a um traço com uma herdabilidade modesta de 0,4. A maior parte (60%) do desempenho superior dos genitores ocorre em virtude de fatores dominante e ambiental que não são transmitidos para a descendência. Se a herdabilidade fosse 1,0, o valor previsto para a descendência então estaria entre os dos genitores. Se a herdabilidade fosse 0,0, o valor previsto para a descendência então estaria na média da população, tendo em vista que toda a variação ocorreria em virtude de fatores não hereditários.

Seleção de traços complexos

Nosso tópico final a respeito da herdabilidade no sentido restrito é a aplicação da seleção a longo prazo para melhorar o desempenho de uma população em relação a um traço complexo. Ao aplicar a seleção, agricultores ao longo dos últimos 10.000 anos transformaram um grande número de espécies de plantas selvagens no notável arranjo de cultivos de frutas, vegetais, cereais e especiarias de que desfrutamos atualmente. De modo semelhante, criadores de animais aplicaram a seleção para domesticar muitas espécies selvagens, transformando lobos em cães, aves selvagens em galinhas e javalis em porcos.

A seleção é um processo por meio do qual apenas os indivíduos com determinadas características contribuem para o *pool* gênico que forma a próxima geração (ver Capítulos 18 e 20). A seleção aplicada pelos seres humanos para melhorar um cultivo ou rebanho é denominada *seleção artificial*, para distingui-la da seleção natural. Vejamos um exemplo de como a seleção artificial atua.

A provitamina A é um precursor na biossíntese da vitamina A, um nutriente importante para olhos saudáveis e um sistema imune com bom funcionamento. Produtos vegetais são uma fonte importante de provitamina A para os seres humanos; entretanto, pessoas em muitas áreas do globo têm muito pouca provitamina A em suas dietas. Para solucionar esse problema, um agricultor busca aumentar o conteúdo de provitamina A de uma população de milho utilizada em partes da América Latina nas quais a deficiência de vitamina A é comum. Atualmente, essa população produz 1,25 µg de provitamina A por grama de grãos de milho. A variância em relação à população é de 0,06 μg^2 (Figura 19.9). Para melhorar a população, o agricultor seleciona um grupo de plantas que produz 1,5 µg ou mais de provitamina A por grama de grãos de milho. A média em relação ao grupo selecionado é de 1,63 µg. O agricultor cruza aleatoriamente as plantas selecionadas e cultiva a descendência para produzir a próxima geração, que apresenta média de 1,44 µg por grama de grãos de milho.

Se a herdabilidade no sentido restrito de um traço não é conhecida antes da realização de um experimento de seleção artificial, os resultados de tais experimentos podem ser utilizados para estimá-la. Aqui está um exemplo com a utilização do caso da provitamina A no milho. Iniciaremos com a equação anterior:

$$\hat{x}_0 = h^2 \bar{x}_p$$

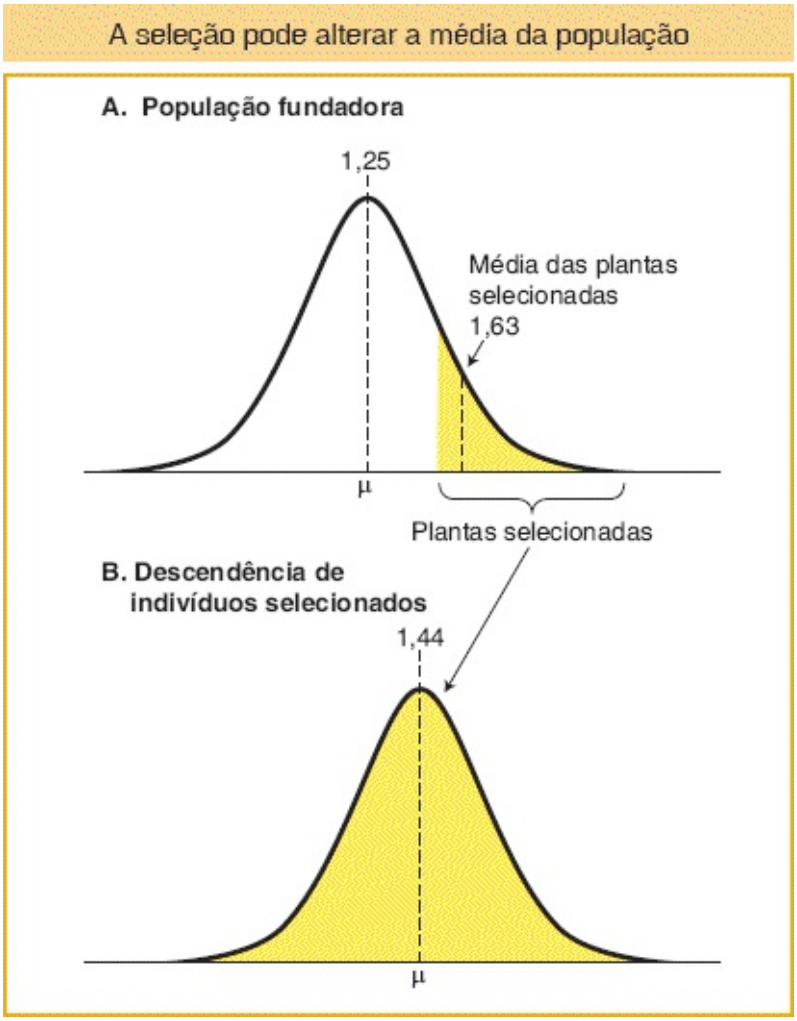


FIGURA 19.9 Distribuição dos valores de traço em relação à provitamina A em grãos de milho em uma população inicial (**A**) e na população de descendentes (**B**) após uma geração de seleção. A população inicial apresentava uma média de 1,25 μg/g, os indivíduos selecionados, uma média de 1,63 μg/g e a população de descendentes, uma média de 1,44 μg/g.

e a reescreveremos como:

$$h^2 = \frac{\bar{x}_o}{\bar{x}_p}$$

$-x_p$ é o desvio médio dos genitores (as plantas selecionadas) da média da população. Esse é conhecido como o **diferencial de seleção (S)**, a diferença entre

a média do grupo selecionado e aquela da população de base. Em relação ao nosso exemplo:

$$-x_p = 1{,}63 - 1{,}25 = 0{,}38$$

$-x_0$ é o desvio médio da descendência da média da população. Essa é conhecida como a **resposta à seleção (R)**, a diferença entre a média da descendência e aquela da população original. Em relação ao nosso exemplo:

$$-x_0 = 1{,}44 - 1{,}25 = 0{,}19$$

Agora podemos calcular a herdabilidade no sentido restrito em relação a esse traço nessa população como:

$$h^2 = \frac{R}{S} = \frac{\overline{x}_o}{\overline{x}_p} = \frac{0{,}19}{0{,}38} = 0{,}5$$

A lógica de base desse cálculo é que a resposta representa a parte herdada ou aditiva do diferencial de seleção.

Ao longo do último século, geneticistas quantitativos conduziram um grande número de experimentos de seleção como esse. Tipicamente, esses experimentos são realizados ao longo de muitas gerações e são denominados estudos de seleção a longo prazo. A cada geração, os melhores indivíduos são selecionados para produzir a geração subsequente. Tais estudos foram realizados em espécies economicamente importantes, tais como plantas de cultivo e criações animais e em muitos organismos-modelo, como *Drosophila*, camundongos e nematódeos. Esse trabalho demonstrou que praticamente qualquer espécie responderá à seleção em relação a praticamente qualquer traço. As populações contêm conjuntos profundos de variação genética aditiva.

Aqui estão dois exemplos de experimentos de seleção a longo prazo. No primeiro experimento, moscas-das-frutas foram selecionadas em relação ao aumento da velocidade de voo ao longo de um período de 100 gerações (Figura 19.10 A). A cada geração, as moscas mais velozes foram selecionadas e cruzadas

para formar a próxima geração. Ao longo das 100 gerações, a velocidade de voo média das moscas na população aumentou de 2 para 170 cm/s e nem as moscas nem os ganhos produzidos pela seleção demonstraram quaisquer sinais de diminuição após 100 gerações. No segundo experimento, os camundongos foram selecionados ao longo de 10 gerações em relação à quantidade de "corrida na roda" que realizavam por dia (Figura 19.10 B). Houve um aumento de 75% ao longo de apenas 10 gerações. Esses estudos e outros similares demonstram o tremendo poder da seleção artificial e de profundos conjuntos de variação genética aditiva nas espécies.

> CONCEITO-CHAVE A herdabilidade no sentido restrito (h^2) é a proporção da variância fenotípica que é atribuível aos efeitos aditivos. Esse tipo de herdabilidade mede até que ponto a variação entre os indivíduos em uma população é transmitida de modo previsível para a sua descendência. O valor de h^2 pode ser estimado de dois modos: (1) com a utilização da correlação entre os genitores e a descendência e (2) com a utilização da razão da resposta à seleção e do diferencial de seleção. O valor de h^2 é importante no cultivo de plantas e na criação de animais, tendo em vista que fornece uma medida do quão bem um traço responderá ao cruzamento seletivo.

19.5 Mapeamento de QTL em populações com heredogramas conhecidos

Os genes que controlam a variação em traços quantitativos (ou complexos) são conhecidos como ***loci* de traço quantitativo** ou **QTL**, abreviadamente. Conforme veremos a seguir, os QTL são genes assim como quaisquer outros a respeito dos quais você aprendeu neste livro. Eles podem codificar enzimas metabólicas, proteínas de superfície celular, enzimas de reparo do DNA, fatores de transcrição ou qualquer uma de muitas outras classes de genes. O que é de interesse aqui é que os QTL apresentam variantes alélicas que tipicamente realizam contribuições

quantitativas relativamente pequenas para o fenótipo.

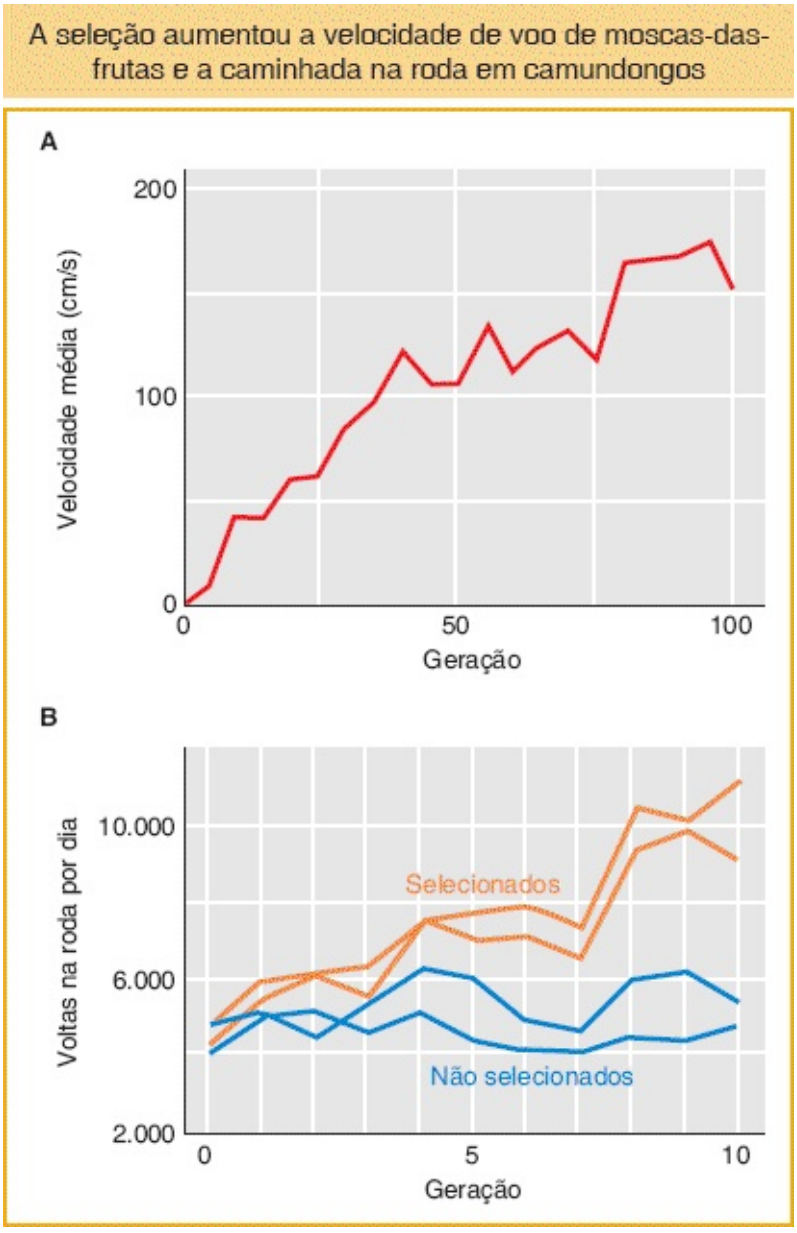


FIGURA 19.10 Resultados de experimentos de seleção a longo prazo. **A.** Seleção para aumento na velocidade de voo de moscas-das-frutas. A velocidade foi testada em um túnel de vento, no qual as moscas voaram contra o vento para alcançar uma fonte de luz. **B.** Seleção de camundongos para aumento na quantidade de caminhada voluntária na roda. (*A.* Dados de K. E. Weber, *Genetics* 144, 1996, 205-213; *B.* Dados de J. G. Swallow *et al.*, *Behav. Genet.* 28, 1998, 227-237.)

Podemos visualizar as contribuições dos alelos em um QTL para o valor do traço ao observar as distribuições de frequências associadas a cada genótipo em

um QTL, conforme demonstrado na Figura 19.11. O *locus* QTL é *B* e as classes genotípicas são *B*/*B*, *B*/*b* e *b*/*b*. Os indivíduos *B*/*B* tendem a apresentar valores de traço mais altos, *B*/*b*, valores intermediários e *b*/*b*, valores pequenos. Entretanto, suas distribuições se sobrepõem e não podemos determinar o genótipo simplesmente observando o fenótipo de um indivíduo, conforme conseguimos fazer em relação aos genes que segregam em razões mendelianas. Na Figura 19.11, um indivíduo com um valor de traço intermediário pode ser *B*/*B*, *B*/*b* ou *b*/*b*.

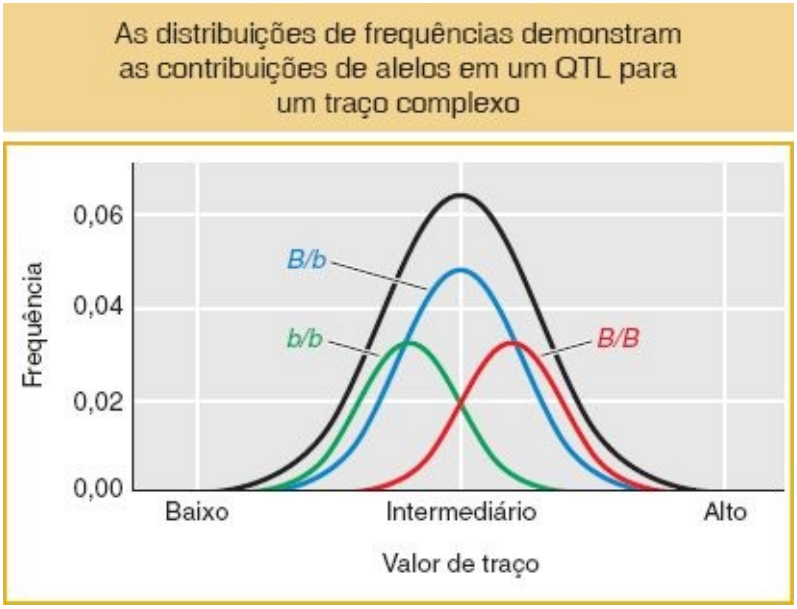


FIGURA 19.11 Distribuições de frequências demonstrando como as distribuições das diferentes classes genotípicas no *locus* B de QTL estão relacionadas com a distribuição geral na população (*linha preta*).

Em virtude dessa propriedade dos QTL, precisamos de ferramentas especiais para determinar sua localização no genoma e caracterizar seus efeitos sobre a variação do traço. Nesta seção, revisaremos um tipo poderoso de análise para a conquista do primeiro desses objetivos. Esse tipo de análise é denominado **mapeamento de QTL**. Ao longo das últimas duas décadas, o mapeamento de QTL revolucionou a nossa compreensão sobre a herança de traços quantitativos. Foi realizado um trabalho pioneiro no mapeamento de QTL com plantas de cultivo tais como tomate e milho. Entretanto, ele tem sido amplamente aplicado em organismos-modelo, tais como camundongo, *Drosophila* e *Arabidopsis*. Mais

recentemente, biólogos evolutivos empregaram o mapeamento de QTL para investigar a herança de traços quantitativos em populações naturais.

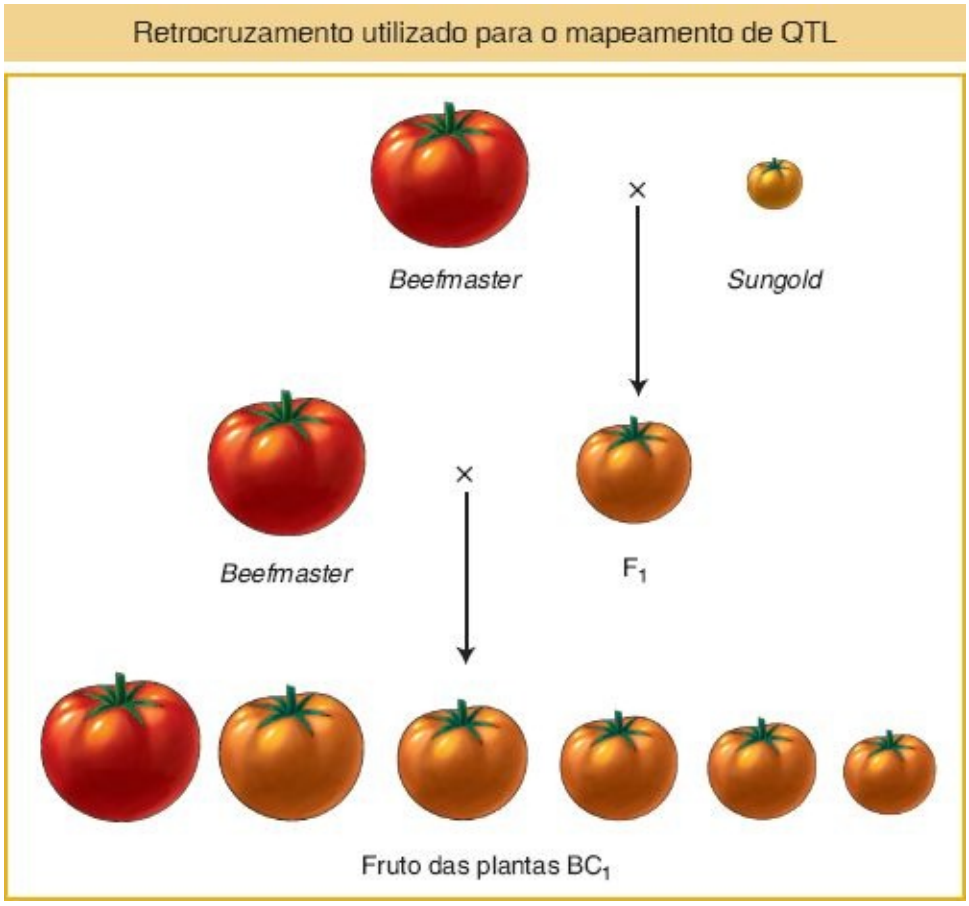


FIGURA 19.12 Esquema de cruzamento em relação a uma população retrocruzada entre tomates *Beefmaster* e *Sungold*. Na geração BC_1, existe uma variação contínua no tamanho dos frutos.

A ideia fundamental por trás do mapeamento de QTL é que é possível identificar a localização de QTL no genoma com a utilização de *loci* marcadores ligados a um QTL. Aqui está como o método funciona. Suponha que você realiza um cruzamento entre duas linhagens endocruzadas – genitor um (P_1) com alto valor de traço e genitor dois (P_2) com um baixo valor de traço. A F_1 pode ser retrocruzada com P_1 para criar uma população BC_1, na qual os alelos em todos os genes nos dois genomas parentais irão segregar. *Loci* marcadores, tais como os SNP ou microssatélites, podem ser pontuados de modo não ambíguo como P_1 homozigoto ou heterozigoto em relação a cada indivíduo de BC_1. Se houver um

QTL ligado ao *locus* marcador, o valor de traço médio dos indivíduos homozigotos P_1 no *locus* marcador será então diferente do valor de traço médio de indivíduos heterozigotos. Com base em tais evidências, podemos inferir que um QTL está localizado próximo do *locus* marcador. Vejamos mais detalhadamente como isso funciona.

Método básico

Existe uma diversidade de desenhos experimentais que podem ser utilizados em experimentos de mapeamento de QTL. Iniciaremos descrevendo um desenho simples. Digamos que temos duas linhagens endocruzadas de tomates que diferem no peso do fruto – *Beefmaster*, com frutos de 230 g de peso, e *Sungold*, com frutos de 10 g de peso (Figura 19.12). Cruzamos as duas linhagens para produzir um híbrido de F_1 e em seguida retrocruzamos a F_1 com a linhagem *Beefmaster* para produzir uma geração BC_1. Cultivamos diversas centenas de plantas de BC_1 até a maturidade e medimos o peso dos frutos em cada uma delas. Também extraímos o DNA de cada uma das plantas de BC_1. Utilizamos essas amostras de DNA para determinar o genótipo de cada planta em um conjunto de *loci* marcadores (SNP ou SSR) que estão distribuídos entre todos os cromossomos, de modo que temos um *locus* marcador a cada 5 a 10 centimorgans.

A partir desse processo, montaríamos um conjunto de dados em relação a diversas centenas de plantas e 100 ou mais *loci* marcadores distribuídos pelo genoma. A Tabela 19.6 demonstra parte de um conjunto de dados em relação a apenas 20 plantas e 5 *loci* marcadores, que estão ligados em um único cromossomo. Em relação a cada planta de BC_1, temos o peso de seus frutos e os genótipos nos *loci* marcadores. Você observará que os valores de traço em relação às plantas de BC_1 são intermediários entre os dois genitores, conforme esperado, mas estão mais próximos do valor de *Beefmaster*, tendo em vista que essa é uma população de BC_1 e *Beefmaster* foi o genitor do retrocruzamento. Além disso, tendo em vista que essa é uma população originada por retrocruzamento, os genótipos em cada *locus* marcador são homozigotos em relação ao alelo *Beefmaster* (*B*/*B*) ou heterozigotos (*B*/*S*). Na Tabela 19.6, você pode observar as posições dos *crossing overs* entre os *loci* marcadores que

ocorreram durante a meiose no genitor F_1 da geração BC_1. Por exemplo, a planta BC_1-001 apresenta um cromossomo recombinante com um *crossing over* entre os *loci* marcadores *M3* e *M4*.

O peso do fruto médio geral em relação à população de BC_1 é 175,7. Também podemos calcular a média em relação às duas classes genotípicas em cada *locus* marcador, conforme demonstrado na Tabela 19.6. Em relação ao marcador *M1*, as médias para as classes genotípicas *B*/*B* (176,3) e *B*/*S* (175,3) estão muito próximas da média geral (175,7). Essa é a expectativa se não houver um QTL que afete o peso do fruto próximo de *M1*. Em relação ao marcador *M3*, as médias para as classes genotípicas *B*/*B* (180,7) e *B*/*S* (169,6) são consideravelmente diferentes da média geral (175,7) e entre si. Essa é a expectativa se houver um QTL que afete o peso do fruto próximo de *M3*. Portanto, temos evidências em relação a um QTL que afeta o peso do fruto próximo do marcador *M3*. Observe também que a classe *B*/*B* apresenta um fruto mais pesado do que a classe *B*/*S* de *M3*. As plantas que herdaram o alelo *S* da linhagem *Sungold* com frutos pequenos apresentam frutos menores do que aquelas que herdaram o alelo *B* da linhagem *Beefmaster*.

Tabela 19.6 Dados simulados de peso do fruto e do *locus* marcador em relação a uma população de retrocruzamento entre duas linhagens endocruzadas de tomates – *Beefmaster* e *Sungold*.

		Marcadores				
Planta	**Peso do fruto (g)**	**M1**	**M2**	**M3**	**M4**	**M5**
Beefmaster	230	B/B	B/B	B/B	B/B	B/B

Sungold	10	S/S	S/S	S/S	S/S	S/S
BC_1-001	183	B/B	B/B	B/B	B/S	B/S
BC_1-002	176	B/S	B/S	B/B	B/B	B/B
BC_1-003	170	B/B	B/S	B/S	B/S	B/S
BC_1-004	185	B/B	B/B	B/B	B/S	B/S
BC_1-005	182	B/B	B/B	B/B	B/B	B/B
BC_1-006	170	B/S	B/S	B/S	B/S	B/B
BC_1-007	170	B/B	B/S	B/S	B/S	B/S
BC_1-008	174	B/S	B/S	B/S	B/S	B/S
BC_1-009	171	B/S	B/S	B/S	B/B	B/B
BC_1-010	180	B/S	B/S	B/B	B/B	B/B
BC_1-011	185	B/S	B/B	B/B	B/S	B/S
BC_1-012	169	B/S	B/S	B/S	B/S	B/S
BC_1-013	165	B/B	B/B	B/S	B/S	B/S

BC_1-014	181	B/S	B/S	B/B	B/B	B/S
BC_1-015	169	B/S	B/S	B/S	B/B	B/B
BC_1-016	182	B/B	B/B	B/B	B/S	B/S
BC_1-017	179	B/S	B/S	B/B	B/B	B/B
BC_1-018	182	B/S	B/B	B/B	B/B	B/B
BC_1-019	168	B/S	B/S	B/S	B/B	B/B
BC_1-020	173	B/B	B/B	B/B	B/B	B/B
Média de B/B	–	176,3	179,6	180,7	176,1	175,0
Média de B/S	–	175,3	173,1	169,6	175,3	176,4
Média geral	175,7					

A Figura 19.13 é uma representação gráfica dos dados de mapeamento de QTL de muitas plantas ao longo de um cromossomo. Os dados fenotípicos em relação às classes genotípicas *B*/*B* e *B*/*S* estão representados como distribuições de frequência, de modo que podemos observar as distribuições dos valores de traço.

No marcador *M1*, as distribuições estão totalmente sobrepostas e as médias em relação às distribuições de *B*/*B* e *B*/*S* estão muito próximas. Aparentemente as classes *B*/*B* e *B*/*S* apresentam a mesma distribuição subjacente. No marcador *M3*, as distribuições estão apenas parcialmente sobrepostas e as médias em relação às distribuições de *B*/*B* e *B*/*S* são consideravelmente diferentes. As classes *B*/*B* e *B*/*S* apresentam diferentes distribuições subjacentes, semelhantes à situação na Figura 19.11. Temos evidência de um QTL próximo de *M3*.

Conforme demonstrado na Figura 19.13, as médias de traço em relação aos grupos *B*/*B* e *B*/*S* em alguns marcadores são quase as mesmas. Em outros marcadores, essas médias são um tanto quanto diferentes. Quão diferentes elas precisam ser antes que declaremos que um QTL está localizado próximo de um marcador? Os detalhes estatísticos para responder a essa questão estão além do escopo deste texto. Entretanto, revisaremos a lógica básica por trás da estatística. A análise estatística envolve o cálculo da probabilidade de observação dos dados (os pesos dos frutos específicos e os genótipos do *locus* marcador em relação a todas as plantas), considerando que exista um QTL próximo do *locus* marcador e a probabilidade de observação dos dados, considerando que *não* exista um QTL próximo do *locus* marcador. A razão dessas duas probabilidades é denominada *odds* (chances):

$$odds = \frac{\text{Prob (dados|QTL)}}{\text{Prob (dados|sem QTL)}}$$

A linha vertical | significa “considerando que” e o termo Prob (dados|QTL) é lido como “a probabilidade de observar os dados, considerando que exista um QTL”. Se a probabilidade dos dados quando houver um QTL for 1,0 e a probabilidade dos dados quando não houver um QTL for 0,001, então as *odds* são de 0,1/0,001 = 100. Ou seja, as chances são de 100 em 1 em favor da existência de um QTL. Pesquisadores reportam o $\log_{10}$ das *odds* ou *Lod score*. Assim, se a razão das chances (*odds ratio*) for 100, então o $\log_{10}$ de 100, ou o *Lod score*, é 2,0.

Se houver um QTL próximo do marcador, os dados então foram coletados de duas distribuições subjacentes – uma distribuição em relação à classe *B*/*B* e uma

em relação à classe *B*/*S*. Cada uma dessas distribuições apresenta a sua própria média e variância. Se não houver um QTL, então os dados foram coletados a partir de uma única distribuição, em relação à qual a média e a variância são aquelas da população de BC_1 inteira. Em um *locus* marcador *M1* na Figura 19.13, as distribuições das classes *B*/*B* e *B*/*S* são quase idênticas. Portanto, existe uma alta probabilidade de que os dados tenham sido coletados a partir de uma única distribuição subjacente. No marcador *M3*, as distribuições das classes *B*/*B* e *B*/*S* são consideravelmente diferentes. Portanto, existe mais alta probabilidade de observação dos nossos dados se inferirmos que as plantas *B*/*B* tenham sido coletadas de uma distribuição e as plantas *B*/*S*, de outra.

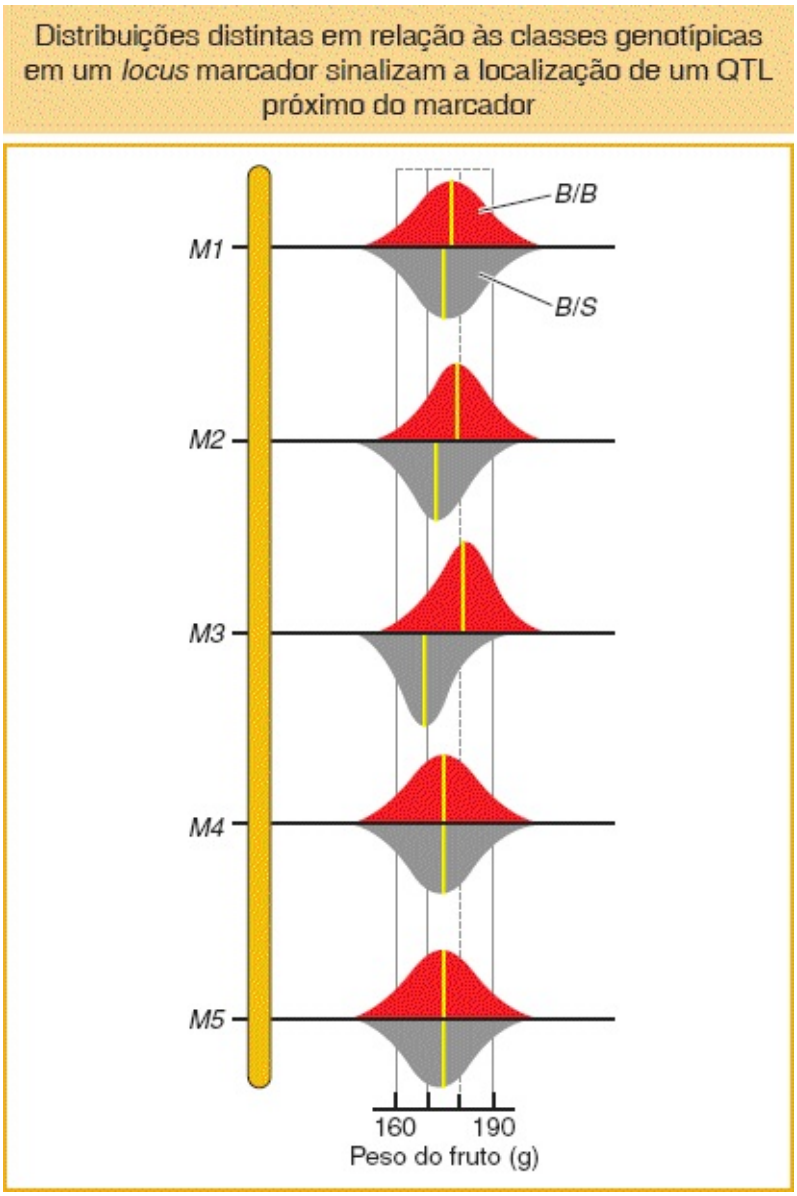


FIGURA 19.13 Um segmento cromossômico de tomate com *loci* marcadores *M1* a *M5*. Em cada *locus* marcador, estão demonstradas as distribuições de frequências em relação ao peso dos frutos de uma população de BC_1 de um cruzamento de *Beefmaster* × *Sungold*.

Além de pesquisar QTL nos *loci* marcadores onde os genótipos são conhecidos, *Lod scores* podem ser calculados em relação a pontos entre os marcadores. Isso pode ser feito com a utilização dos genótipos dos marcadores flanqueadores para inferir os genótipos em pontos entre os marcadores. Por exemplo, na Tabela 19.6, a planta BC_1-001 é *B*/*B* nos marcadores *M1* e *M2* e, assim, ela apresenta uma alta probabilidade de ser *B*/*B* em todos os pontos entre

eles. A planta BC_1-003 é *B*/*B* no marcador *M1*, mas *B*/*S* em *M2* e, assim, a planta pode ser *B*/*B* ou *B*/*S* nos pontos entre eles. A equação da *odds* incorpora essa incerteza quando o *Lod score* é calculado em pontos entre os marcadores.

Os *Lod scores* podem ser plotados em gráficos ao longo do cromossomo, conforme demonstrado pela linha azul na Figura 19.14. Tais gráficos tipicamente demonstram alguns picos de altura variável, bem como trechos que são relativamente planos. Os picos representam QTL putativos, mas quão alto um pico precisa ser antes que declaremos que ele representa um QTL? Conforme discutido nos Capítulos 4 e 18, podemos estabelecer um limiar estatístico para rejeitar a "hipótese nula". Nesse caso, a hipótese nula é que "*não* existe um QTL em uma posição específica ao longo do cromossomo". Quanto maior o *Lod score*, mais baixa então a probabilidade sob a hipótese nula. Existem diferentes procedimentos estatísticos para o estabelecimento de um "valor limiar" para o *Lod score*. Onde o *Lod score* excede o valor limiar, rejeitamos então a hipótese nula, em favor da hipótese alternativa de que um QTL está localizado naquela posição. Na Figura 19.14, o *Lod score* excede o valor limiar (*linha vermelha*) próximo do *locus* marcador *M3*. Concluímos que um QTL está localizado próximo de *M3*.

Além de populações oriundas de retrocruzamento, o mapeamento de QTL pode ser realizado com populações F_2 e outros desenhos de cruzamento. Uma vantagem da utilização de uma população F_2 é que são obtidas estimativas dos valores de traço médios em relação a todos os três genótipos de QTL: genitor homozigoto-1, genitor homozigoto-2 e heterozigoto. Com esses dados, podemos obter estimativas dos efeitos aditivos (A) e da dominância (D) do QTL, conforme discutido anteriormente neste capítulo. Portanto, o mapeamento de QTL possibilita que aprendamos sobre a ação gênica, seja dominante ou aditiva, em relação a cada QTL.

Aqui está um exemplo. Suponha que tenhamos estudado uma população F_2 de um cruzamento de tomates *Beefmaster* e *Sungold* e que tenhamos identificado dois QTL para o peso do fruto. Os pesos médios dos frutos das diferentes classes genotípicas no QTL pode ser semelhante a algo como:

	Peso dos frutos			Efeitos	
	B*/*B	***B*/*S***	***S*/*S***	**A**	**D**
QTL 1	180	170	160	10	0
QTL 2	200	185	110	45	30

Podemos utilizar esses valores do peso dos frutos para o QTL para calcular os efeitos aditivo e de dominância. O QTL 1 é puramente aditivo (D = 0), mas o QTL 2 apresenta um grande efeito de dominância. Além disso, observe que o efeito aditivo do QTL 2 é mais do que 4 vezes aquele do QTL 1 (45 *versus* 10). Alguns QTL apresentam grandes efeitos e outros apresentam efeitos pequenos.

O que pode ser aprendido a partir do mapeamento de QTL? Com os desenhos de mapeamento de QTL mais poderosos, os geneticistas podem estimar (1) o número de QTL (genes) que afetam um traço, (2) as localizações genômicas desses genes, (3) o tamanho dos efeitos de cada QTL, (4) o modo de ação gênica do QTL (dominante *versus* aditiva) e (5) se um QTL afeta a ação de outro QTL (interação epistática). Em outras palavras, podemos obter uma descrição um tanto quanto completa da arquitetura genética do traço.

Muito foi aprendido a respeito da arquitetura genética a partir dos estudos de mapeamento de QTL em diversos organismos. Aqui estão dois exemplos. Primeiramente, o tempo de floração no milho é um traço quantitativo ou contínuo clássico, e de importância crítica no cultivo do milho, tendo em vista que as plantas precisam florescer e amadurecer antes do final da temporada de crescimento. O milho do Canadá está adaptado para florescer dentro de 45 dias após o plantio, enquanto o milho do México pode necessitar de 120 dias ou mais. O mapeamento de QTL demonstrou que a arquitetura genética em relação ao tempo de florescimento no milho envolve mais de 50 genes. Os resultados de um experimento estão demonstrados na Figura 19.15 A; esses resultados mostram

evidências em relação a 15 QTL. Os QTL de tempo de floração do milho em geral apresentam um efeito pequeno, de tal modo que a substituição de um alelo por outro em um QTL altera o tempo de floração em apenas 1 dia ou menos. Portanto, a diferença no tempo de floração entre o milho tropical e o de regiões temperadas envolve muitos QTL.

Em segundo lugar, camundongos têm sido utilizados para mapear QTL de muitos traços de suscetibilidade a doenças. O que é aprendido a respeito dos genes de suscetibilidade a doenças em camundongos com frequência é verdadeiro também em seres humanos. A Figura 19.15 B demonstra os resultados de uma varredura genômica em camundongos para QTL de densidade mineral óssea (DMO), o traço subjacente à osteoporose. Essa varredura identificou dois QTL, um no cromossomo 9 e um no cromossomo 12. A partir de estudos como esse, pesquisadores identificaram mais de 80 QTL em camundongos que podem contribuir para a suscetibilidade à osteoporose. Foram realizados estudos semelhantes em dúzias de outras doenças.

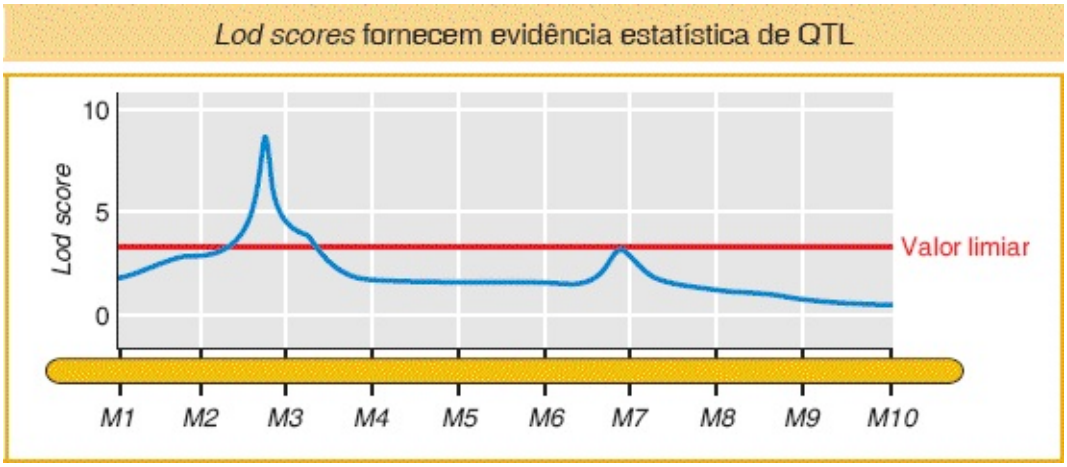


FIGURA 19.14 Gráfico dos *Lod score* de um experimento de mapeamento de QTL ao longo de um cromossomo com 10 *loci* marcadores. A *linha azul* demonstra o valor do *Lod score* em cada posição. Onde o *Lod score* excede o valor limiar, existe uma evidência estatística de um QTL.

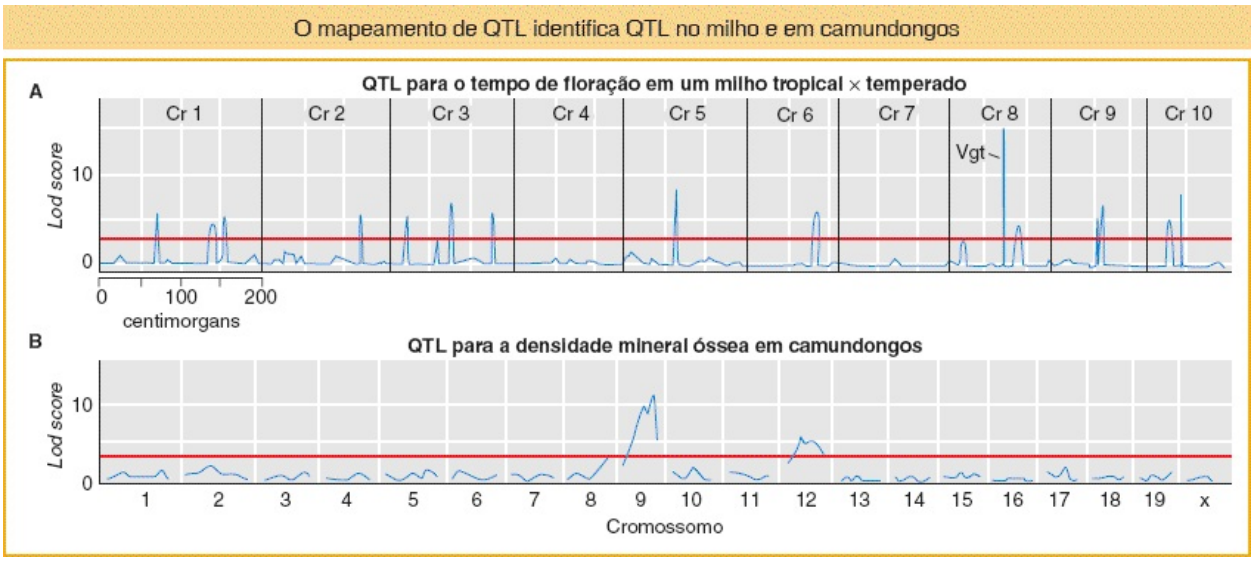


FIGURA 19.15 Gráfico de *Lod scores* de varreduras genômicas para QTL. **A.** Resultados de uma varredura para QTL do tempo de floração no milho. **B.** Resultados de uma varredura para QTL da densidade mineral óssea em camundongos. (*A.* Dados de E. S. Buckler *et al.*, *Science* 325, 2009, 714-718; *B.* Dados de N. Ishimori *et al.*, *J. Bone Min. Res.* 23, 2008, 1529-1537.)

Do QTL ao gene

O mapeamento de QTL tipicamente não revela a identidade do(s) gene(s) no QTL. No melhor caso, a resolução do mapeamento de QTL está na ordem de 1 a 10 cM, o tamanho de uma região que pode conter 100 ou mais genes. O direcionamento do QTL para um gene único requer experimentos adicionais para o **mapeamento fino** de um QTL. Para tanto, o pesquisador cria um conjunto de estoques genéticos homozigotos (também denominados *linhagens*), cada um com um *crossing over* próximo do QTL. Esses estoques ou linhagens diferem entre si próximo do QTL, mas são idênticos uns aos outros (**isogênicos**) em todo o restante de seus genomas. As linhagens que são idênticas em seus genomas inteiros, com exceção de uma pequena região de interesse, são denominadas **linhagens congênicas** ou **quase isogênicas**. O isolamento de QTL em um contexto isogênico é crítico, tem em vista que apenas a única região do QTL difere entre as linhagens congênicas. Portanto, a utilização de linhagens congênicas elimina as complicações causadas pela presença de múltiplos QTL que segregam ao mesmo tempo.

Com a utilização do exemplo anterior do peso do fruto do tomate, a região cromossômica de um conjunto das referidas linhagens congênicas está

demonstrada na Figura 19.16. Os genes (*flc*, *arf4*,…) estão demonstrados na parte superior e a localização de cada *crossing over* está indicada pela alteração da cor, do vermelho (genótipo *Beefmaster*) para o amarelo (genótipo *Sungold*). O peso médio do fruto em relação às linhagens congênicas que carregam esses cromossomos recombinantes está indicado à direita. Ao inspecionar a Figura 19.16, você observará que todas as linhagens com alelo *Beefmaster* de *kin1* (um gene de quinase) apresentam frutos de aproximadamente 180 g, enquanto aquelas com o alelo *Sungold* de *kin1* apresentam frutos de aproximadamente 170 g. Nenhum dos outros genes está associado ao peso dos frutos dessa maneira. Se confirmado por meio de testes estatísticos apropriados, esse resultado possibilita que identifiquemos o *kin1* como o gene subjacente a esse QTL.

A Tabela 19.7 lista uma pequena amostra de centenas de genes ou QTL que afetam a variação quantitativa de diferentes espécies que foram identificados. A lista inclui o gene em relação ao tempo de floração do milho, *Vgt*, que é a base de um dos picos de *Lod* na Figura 19.15 A. Um aspecto notável dessa lista é a diversidade de funções gênicas. Aparentemente não existe uma regra no sentido de que apenas tipos de genes em particular possam ser um QTL. A maior parte dos genes nos genomas dos organismos, se não todos, provavelmente contribui para uma variação quantitativa nas populações.

> CONCEITO-CHAVE O mapeamento de *locus* de traço quantitativo (QTL) é um procedimento para a identificação das localizações genômicas dos genes (QTL) que controlam a variação em relação aos traços quantitativos ou complexos. O mapeamento de QTL avalia a progênie de cruzamentos controlados em relação aos seus genótipos em marcadores moleculares e em relação aos seus valores de traço. Se os diferentes genótipos em um *locus* marcador apresentarem diferentes valores médios para o traço, existe então evidência de um QTL próximo do marcador. Após uma região do genoma que contém um QTL ter sido identificada, o QTL pode ser mapeado até genes únicos com a utilização de linhagens congênicas.

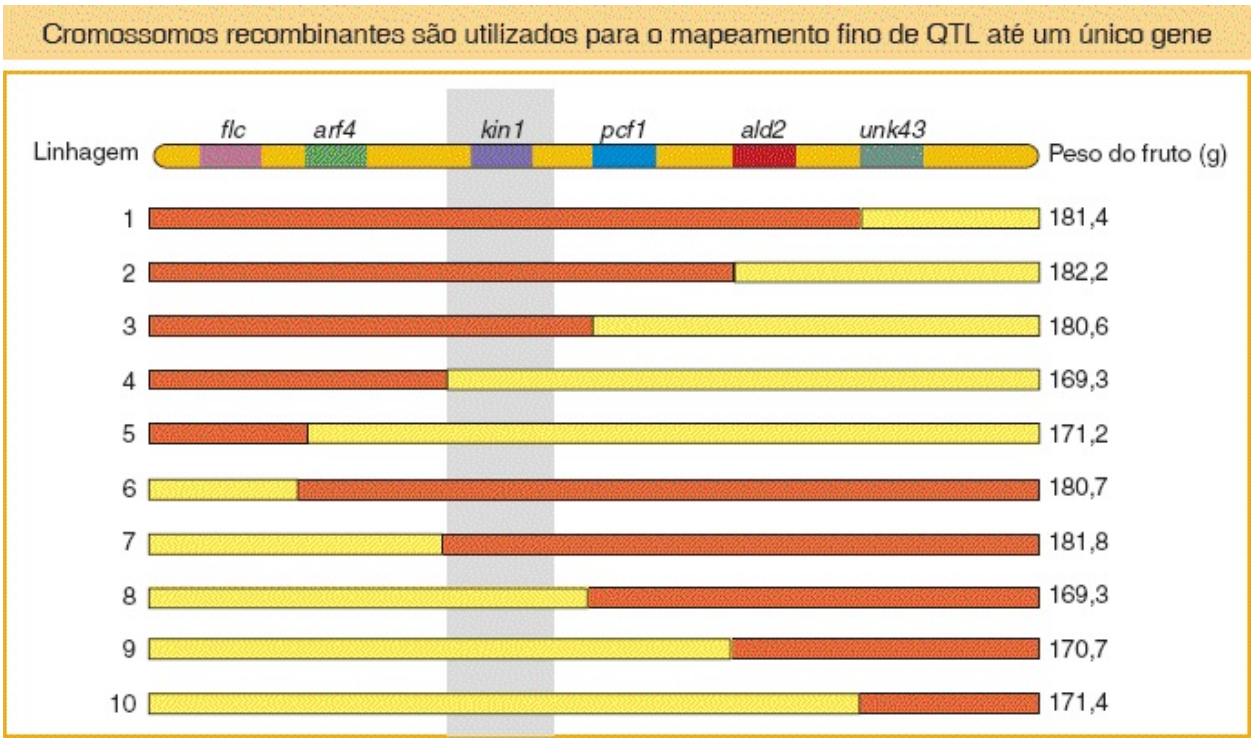


FIGURA 19.16 Um segmento cromossômico de tomate para um conjunto de 10 linhagens congênicas que apresentam *crossing overs* próximos de um QTL para peso do fruto. Os segmentos cromossômicos *vermelhos* são derivados da linhagem *Beefmaster* e os segmentos *amarelos*, da linhagem *Sungold*. Diferenças no peso do fruto entre as linhagens tornam possível a identificação do gene *kin1* como o gene subjacente a esse QTL.

Tabela 19.7 Alguns genes que contribuem para a variação quantitativa e que foram identificados pela primeira vez com a utilização do mapeamento de QTL.

Organismo	Traço	Gene	Função gênica
Levedura	Crescimento em alta temperatura	*RHO2*	GTPase
Arabidopsis	Tempo para floração	*CRY2*	Criptocromo
Milho	Ramificação	*Tb1*	Fator de transcrição

Milho	Tempo para floração	*Vgt*	Fator de transcrição
Arroz	Sensibilidade ao fotoperíodo	*Hd1*	Fator de transcrição
Arroz	Sensibilidade ao fotoperíodo	*CK2*α	Subunidade α da caseinoquinase
Tomate	Conteúdo de açúcar do fruto	*Brix9-2-5*	Invertase
Tomate	Peso do fruto	*Fw2.2*	Sinalização entre células
Drosophila	Número de cerdas	*Scabrous*	Glicoproteína secretada
Vaca	Produção leiteira	*DGAT1*	Diacilglicerol acetiltransferase
Camundongos	Câncer de cólon	*Mom1*	Modificador de um gene supressor de tumor
Camundongos	Diabetes melito do tipo 1	*I-A*β	Antígeno de histocompatibilidade
			Proteína contendo

Seres humanos	Asma	*ADAM33*	domínio de metaloproteinase
Seres humanos	Doença de Alzheimer	*ApoE*	Apolipoproteína
Seres humanos	Diabetes melito do tipo 1	*HLA-DQA*	Glicoproteína de superfície de MHC de classe II

Fonte: A. M. Glazier *et al.*, *Science* 298, 2002, 2345-2349.

19.6 Mapeamento de associação em populações de cruzamento aleatório

Se você leu recentemente alguma notícia anunciando que pesquisadores identificaram um gene de suscetibilidade para autismo, diabetes melito, hipertensão arterial ou algum outro distúrbio, existe uma excelente chance de que o gene tenha sido descoberto com a utilização da técnica que iremos revisar, denominada **mapeamento de associação**. O mapeamento de associação é um método para encontrar QTL no genoma com base no desequilíbrio de ligação de ocorrência natural (ver Capítulo 18) entre um *locus* marcador e o QTL em uma população de cruzamento aleatório. Tendo em vista que ele utiliza o desequilíbrio de ligação, o método também é denominado mapeamento de desequilíbrio de ligação. Conforme veremos, esse método com frequência possibilita que os pesquisadores identifiquem diretamente os genes específicos que controlam as diferenças no fenótipo entre os membros de uma população.

A ideia básica por trás do mapeamento de associação tem sido apresentada e utilizada durante décadas. Aqui está um exemplo da década de 1990 em relação ao gene *ApoE* em seres humanos, um gene envolvido no metabolismo de

lipoproteínas (complexos de lipídios e proteínas). Em virtude do seu papel no metabolismo das lipoproteínas, o *ApoE* foi considerado um **gene candidato** para um papel causal nas doenças cardiovasculares, o acúmulo de depósitos de gordura (lipídio) nas artérias. Pesquisadores procuraram por associações estatísticas entre os alelos do *ApoE* que as pessoas carreiam e se elas apresentavam doença cardiovascular. Eles encontraram uma associação entre o alelo *e4* desse gene e a doença – pessoas que carreavam o alelo *e4* apresentavam 42% mais probabilidade de ter a doença do que aquelas que carreavam outros alelos. Embora esse tipo de estudo tenha obtido sucesso, ele exigia que um gene candidato, que se suspeitava afetar o traço, fosse conhecido previamente.

Ao longo da última década, os avanços nas tecnologias genômicas catalisaram a aplicação em ampla escala do mapeamento de associação. Em particular, o mapeamento de associação foi revolucionado por meio do desenvolvimento de mapas de SNP do genoma inteiro e das tecnologias de genotipagem de alta produção, que possibilitam a identificação de centenas de milhares de SNP em dezenas de milhares de indivíduos (ver Capítulo 18). O mapeamento de associação atualmente é utilizado de modo rotineiro para realizar a varredura do genoma inteiro em relação aos genes que contribuem para a variação quantitativa. Esse tipo de estudo é conhecido como estudo de **associação ampla do genoma** (estudo de **GWA** ou **GWAS**). Uma vantagem importante dos estudos de GWA é que os genes candidatos não são necessários, uma vez que é realizada a varredura de todos os genes no genoma.

O mapeamento de associação oferece diversas vantagens sobre o mapeamento de QTL. Primeiramente, tendo em vista que ele é realizado com populações de cruzamento aleatório, não há necessidade de realizar cruzamentos controlados ou de trabalhar com famílias humanas com relações de genitores e descendência conhecidas. Em segundo lugar, ele testa muitos alelos em um *locus* de uma vez. Nos estudos de mapeamento de QTL, existem dois genitores (tomates *Beefmaster* e *Sungold* no exemplo anterior) e, assim, apenas dois alelos estão sendo comparados. Com o mapeamento de associação, todos os alelos na população estão sendo analisados ao mesmo tempo. Finalmente, o mapeamento de associação pode levar à identificação direta dos genes no QTL sem a necessidade

de estudos de mapeamento fino subsequentes. Isso é possível porque os SNP em qualquer gene que influencia o traço demonstrarão associações mais fortes com o traço do que os SNP em outros genes. Vejamos como isso funciona.

Método básico

Iniciaremos observando como a variação genética é padronizada no genoma em uma população. No Capítulo 18, discutimos o desequilíbrio de ligação (LD) ou a associação não aleatória de alelos em dois *loci*. A Figura 19.17 demonstra como o LD poderia aparecer entre uma amostra de cromossomos de 18 indivíduos diferentes. Os SNP (ou outros polimorfismos) que estão próximos entre si tendem a estar em forte desequilíbrio, enquanto aqueles que estão mais distantes estão em desequilíbrio fraco ou nenhum desequilíbrio. Os genomas também tendem a apresentar *hotspots* de recombinação, pontos nos quais o *crossing over* ocorre a uma alta frequência. Os *hotspots* rompem o desequilíbrio de ligação de tal modo que os SNP em qualquer lado do *hotspot* estão em equilíbrio entre si. Os SNP que não estão separados por um *hotspot* formam um bloco de haplótipo de SNPs fortemente correlacionados.

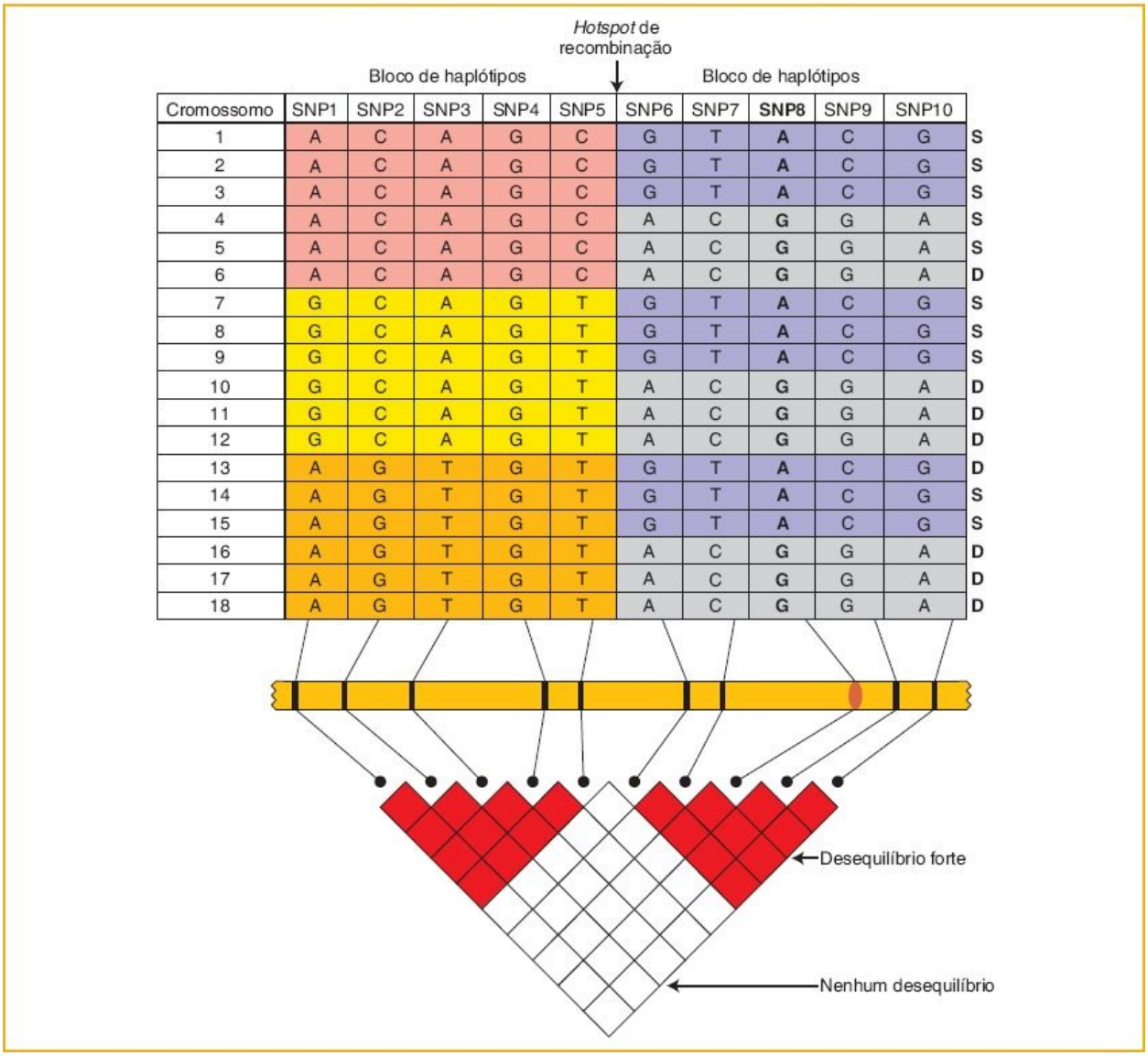
Suponha que o SNP8 na Figura 19.17 seja um SNP em um gene que causa uma diferença no fenótipo, de tal modo que os indivíduos com o genótipo A/A apresentam um fenótipo diferente daqueles com A/G ou G/G. O SNP8 poderia afetar o fenótipo ao causar uma troca de aminoácido ou afetar a expressão gênica. O SNP8 ou quaisquer SNP que afetem diretamente um fenótipo são denominados SNP funcionais. Tendo em vista que o SNP8 está em um forte desequilíbrio com outros SNP no bloco (SNP 6, 7, 9 e 10), qualquer um desses outros SNP pode atuar como um substituto do SNP8 funcional. Os indivíduos que são T/T no SNP7 apresentarão o mesmo fenótipo daqueles que são A/A no SNP8, tendo em vista que o SNP7 e o SNP8 estão em LD. Quando os genótipos de SNP estão correlacionados (em desequilíbrio), o valor de traço estará então correlacionado. Por esse motivo, os estudos de GWA não precisam analisar os SNP funcionais reais, mas precisam apresentar os SNP em cada bloco de haplótipo.

Para conduzir um estudo de GWA em relação a uma condição de doença em seres humanos, podemos analisar 2.000 indivíduos com um distúrbio, tal como o

diabetes melito com início em adultos ou tipo 2. Também poderíamos selecionar outros 2.000 indivíduos de controle que não apresentam esse distúrbio. Cada um dos 4.000 participantes doaria sangue, a partir do qual o seu DNA seria extraído. As amostras de DNA seriam genotipadas em relação a um conjunto de 300.000 SNP que estão distribuídos por todo o genoma. Desejamos um número de SNP suficiente, de tal modo que cada um dos blocos de haplótipo no genoma seja marcado por um ou mais SNP (ver Figura 19.17). O conjunto de dados resultante seria enorme – composto por 300.000 genótipos em 4.000 indivíduos – um total de 1,2 bilhão de pontos de dados. Uma pequena parte de tal conjunto de dados está demonstrada na Tabela 19.8.

Após a montagem dos dados, o pesquisador realiza um teste estatístico em cada SNP para determinar se um de seus alelos está mais frequentemente associado ao diabetes do que o esperado ao acaso. No caso de um traço categórico, tal como ser "afetado" ou "não afetado" pelo diabetes, podem ser utilizados testes estatísticos semelhantes ao teste do χ^2 (ver Capítulo 3). Um teste estatístico é realizado em separado em cada SNP e os valores P são inseridos em um gráfico ao longo do cromossomo. A hipótese nula é que o SNP não está associado ao traço. Se o valor P em relação a um SNP for inferior a 0,05, então a evidência em relação à hipótese nula é fraca e favoreceremos a hipótese alternativa de que os diferentes genótipos no SNP estão associados a diferentes fenótipos do traço. O mapeamento de associação realmente não comprova que um gene ou um SNP em um gene afete um traço. Ele apenas fornece evidências estatísticas em relação a uma associação entre o SNP e o traço. A comprovação requer a caracterização molecular do gene e de seus diferentes alelos.

Hotspots de recombinação rompem o desequilíbrio de ligação



Cromossomo	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	**SNP8**	SNP9	SNP10	
1	A	C	A	G	C	G	T	**A**	C	G	S
2	A	C	A	G	C	G	T	**A**	C	G	S
3	A	C	A	G	C	G	T	**A**	C	G	S
4	A	C	A	G	C	A	C	**G**	G	A	S
5	A	C	A	G	C	A	C	**G**	G	A	S
6	A	C	A	G	C	A	C	**G**	G	A	D
7	G	C	A	G	T	G	T	**A**	C	G	S
8	G	C	A	G	T	G	T	**A**	C	G	S
9	G	C	A	G	T	G	T	**A**	C	G	S
10	G	C	A	G	T	A	C	**G**	G	A	D
11	G	C	A	G	T	A	C	**G**	G	A	D
12	G	C	A	G	T	A	C	**G**	G	A	D
13	A	G	T	G	T	G	T	**A**	C	G	D
14	A	G	T	G	T	G	T	**A**	C	G	S
15	A	G	T	G	T	G	T	**A**	C	G	S
16	A	G	T	G	T	A	C	**G**	G	A	D
17	A	G	T	G	T	A	C	**G**	G	A	D
18	A	G	T	G	T	A	C	**G**	G	A	D

FIGURA 19.17 (*Parte superior*) Diagrama da distribuição de SNP e haplótipos em relação a um segmento cromossômico de 18 indivíduos. Os haplótipos com frequência ocorrem em blocos (regiões de recombinação mais baixa), separados uns dos outros por *hotspots* de recombinação (*diferentes cores indicam os blocos de haplótipos*). (A coluna de S e D à direta são para o Problema 19.4.) O SNP8 (*negrito*) controla uma diferença nos valores de traço. (*Parte inferior*) Você consegue dizer se dois SNP demonstram desequilíbrio ao observar a cor do quadrado onde as linhas dos marcadores apresentam intersecção. Em um bloco de haplótipo, os SNP demonstram forte desequilíbrio. Os SNP em diferentes blocos de haplótipos demonstram desequilíbrio fraco ou nenhum desequilíbrio. (Dados de David Altshuler *et al.*, *Science* 322, 2008, 881-888.)

A Figura 19.18 A demonstra os resultados de um estudo de mapeamento de associação em relação ao tamanho corporal em cães. Cada ponto inserido em

gráfico ao longo dos cromossomos (eixo *x*) representa o valor P (eixo *y*) em relação a um teste de associação entre o tamanho corporal e um SNP. Os valores P são inseridos em gráfico com a utilização de uma escala inversa, de tal modo que quanto mais alto no eixo *x*, menor o valor. No cromossomo 15, existe um aglomerado de SNP acima da linha do limiar, indicando que a hipótese nula de nenhuma associação pode ser rejeitada em relação a esses SNP, em favor da hipótese alternativa de que um gene que afeta o tamanho corporal em cães está localizado nessa posição. O pico forte no cromossomo 15 envolve SNP no gene do *fator de crescimento 1 semelhante à insulina* (*IGF1*), um gene que codifica um hormônio envolvido no crescimento juvenil em mamíferos. Esse gene é o principal contribuinte para a diferença no tamanho entre raças pequenas e grandes de cães (Figura 19.18 B).

Tabela 19.8 Parte de um conjunto de dados simulados para um experimento de mapeamento de associação.

Indivíduo	SNP1	SNP2	SNP3	Diabetes tipo 2	Altura (cm)
1	C/C	A/G	T/T	sim	173
2	C/C	A/A	C/C	sim	170
3	C/G	G/G	T/T	não	183
4	C/G	G/G	C/T	não	180
5	C/C	G/G	C/T	não	173
6	G/G	A/G	C/T	sim	178

7	G/G	A/G	C/T	não	163
8	C/G	G/G	C/T	não	168
9	C/G	A/G	C/T	sim	165
10	G/G	A/A	C/C	sim	157

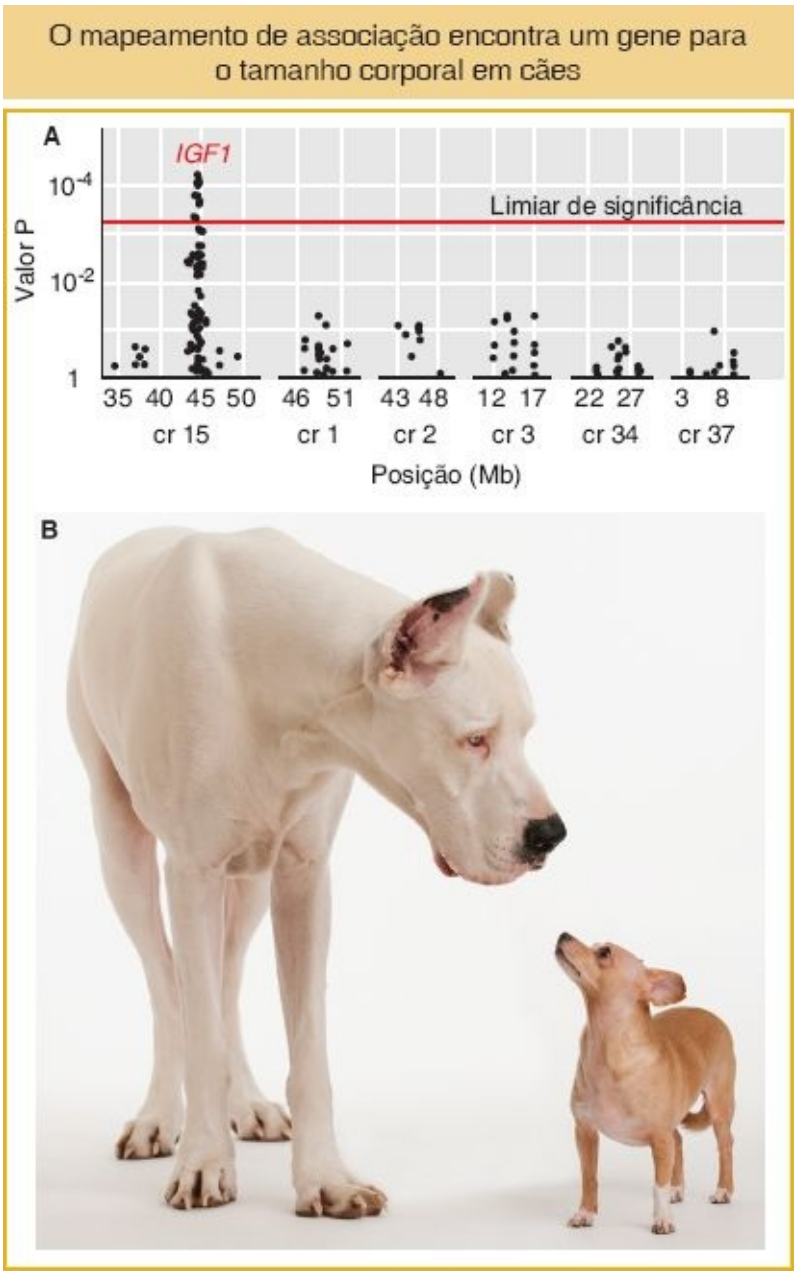


FIGURA 19.18 A. Resultados de um experimento de mapeamento de associação em relação ao tamanho corporal em cães. Cada ponto no gráfico representa o valor P em um teste de associação entre um SNP e o tamanho corporal. Os pontos acima da "linha limiar" demonstram evidência de uma associação estatisticamente significativa. **B.** Exemplos de uma raça pequena e uma grande de cães. (*B* © Tetra Images/Corbis.)

GWA, genes, doença e herdabilidade

Ao longo dos últimos 10 anos, foi realizado um grande número de estudos de

GWA e muito foi aprendido a partir deles a respeito da variação herdada em seres humanos e em outras espécies. Vejamos um dos maiores estudos, que foi uma pesquisa em relação aos genes de risco de doenças em um grupo de 17.000 pessoas, com a utilização de 500.000 SNP. A Figura 19.19 demonstra gráficos dos valores P em relação às associações entre os SNP e diversas doenças comuns. Os pontos verdes são as associações estatisticamente significativas. Note a espícula de pontos verdes no cromossomo 6 para artrite reumatoide e diabetes tipo 1 (juvenil). Essas são duas doenças autoimunes e a espícula está posicionada em um gene do antígeno leucocitário humano (*HLA*) do complexo principal de histocompatibilidade (MHC) de genes que regulam a resposta imune em seres humanos e em outros vertebrados. Portanto, os genes ativos na resposta imune normal estão implicados como uma causa de doenças autoimunes. O gene *PTPN22* também está associado ao risco de diabetes tipo 1. O *PTPN22* codifica a proteína tirosina fosfatase, que é expressa em células linfoides do sistema imune. Em relação à doença coronariana, existe uma associação significativa com o gene *ApoE*, confirmando um estudo inicial mencionado anteriormente.

Estudos de GWA identificaram mais de 300 genes de risco em relação a aproximadamente 70 doenças e os números estão aumentando. Esses dados estão conduzindo a uma nova era da **genômica pessoal**, na qual um indivíduo pode ter seu genoma varrido para determinar o seu genótipo em genes que sabidamente aumentam o risco de doenças. Embora essa ciência seja relativamente jovem, é possível identificar indivíduos que apresentam um risco 10 vezes mais alto em relação a determinadas doenças do que outros membros da população. Tais informações podem ser utilizadas para iniciar medidas de prevenção e alterações no estilo de vida (ambiente) que contribuem para o risco de doenças. Algumas empresas estão propondo oferecer para a compra em sua farmácia local "*kits* de teste genético" em relação a doenças específicas, tal como a doença de Alzheimer. Bioeticistas expressaram a sua preocupação de que os consumidores não estejam preparados para avaliar adequadamente os resultados sem o aconselhamento de profissionais médicos.

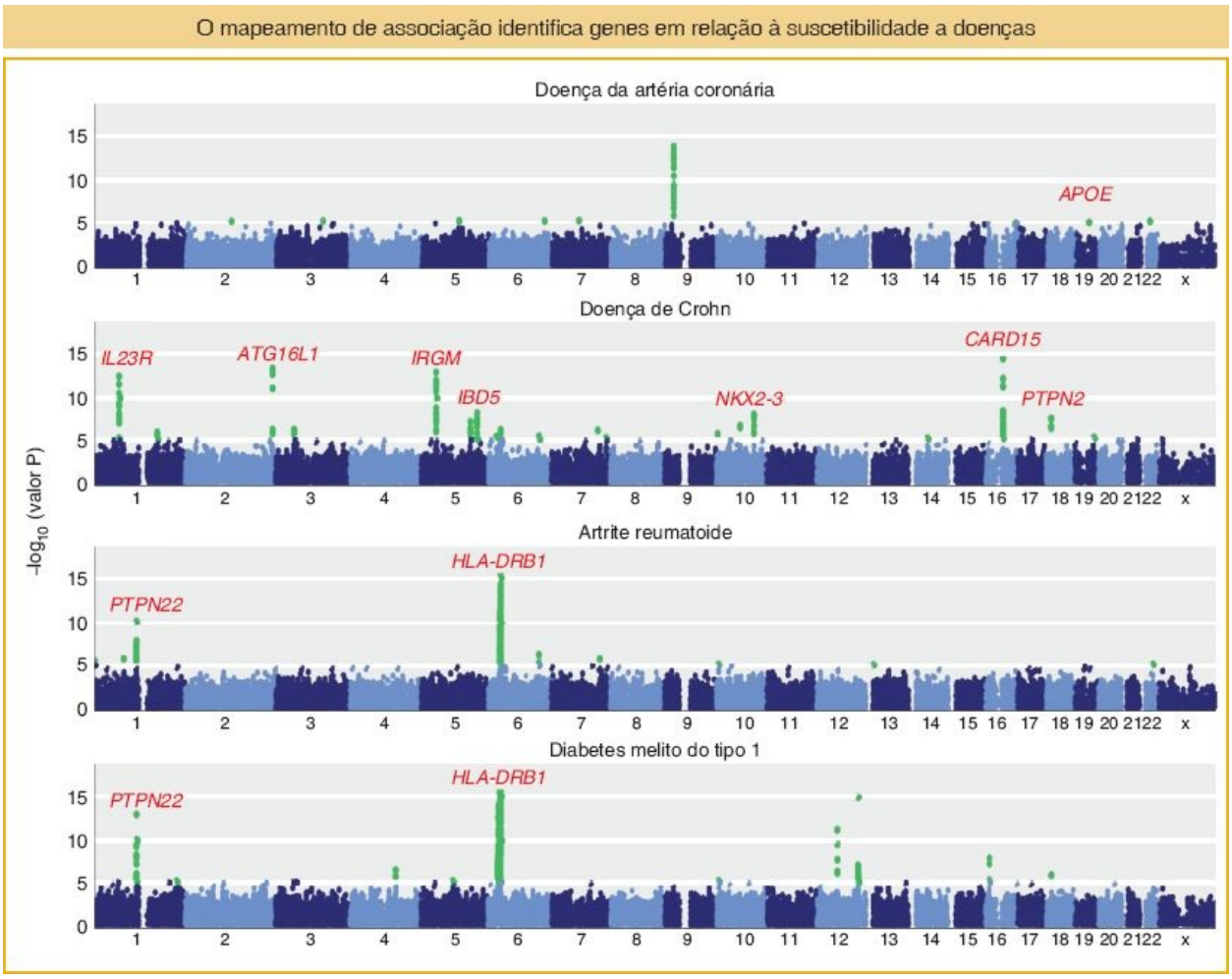


FIGURA 19.19 Resultados de um estudo de associação de genoma inteiro de doenças comuns em seres humanos. Os 23 cromossomos humanos estão dispostos da esquerda para a direita. O eixo *y* demonstra o valor P em relação ao teste estatístico de uma associação entre a doença e cada SNP. Os resultados de testes significativos estão demonstrados como *pontos verdes*. As denominações de alguns genes identificados por meio desta análise estão demonstradas em *vermelho*. (Dados de The Wellcome Trust Case Control Consortium, *Nature* 447, 2007, 661-678.)

Tendo em vista que a altura em seres humanos é um traço quantitativo clássico, geneticistas quantitativos apresentam grande interesse na realização de estudos de GWA em relação a esse traço. Estudos de GWA identificaram mais de 180 genes que afetam a altura. Cada um desses genes apresenta um pequeno efeito aditivo (aproximadamente 1 a 4 mm), conforme esperado em relação a um traço regulado por muitos genes. Entretanto, um resultado desconcertante foi que os 180 genes foram responsáveis por apenas 10% da variância genética na altura. Isso está

muito longe do valor de quase 80% em relação à herdabilidade no sentido amplo para a altura. A diferença entre os 10% e os 80% foi atribuída à perda de herdabilidade. Em relação ao risco de doenças, também existe muita perda de herdabilidade. Por exemplo, estudos de GWA obtiveram sucesso ao explicar apenas 10% da variação genética em relação à doença de Crohn e apenas 5% da variação genética em relação ao diabetes melito do tipo 2.

Foi uma surpresa para muitos geneticistas que estudos de GWA com centenas de milhares de SNP englobando o genoma e amostras de mais de 10.000 indivíduos conseguissem explicar apenas uma pequena fração da variação hereditária. Até o momento, não se sabe a razão disso. Pesquisadores esperavam que doenças comuns, tal como o diabetes melito do tipo 2, fossem causadas por alelos comuns; ou seja, alelos com frequências entre 5 e 95%. Estudos de GWA são projetados para detectar os efeitos de alelos comuns, mas não para detectar os efeitos de alelos raros. Portanto, uma hipótese é que a suscetibilidade em relação a muitas doenças comuns (ou a variação na altura) seja causada por uma grande quantidade de alelos raros. Em outras palavras, os alelos de suscetibilidade a doenças que estão segregados em uma família são diferentes daqueles de outra família não relacionada.

Apesar da incapacidade dos estudos de GWA de explicar toda a variação hereditária em relação aos traços, essa abordagem proporcionou um importante avanço na compreensão da variação genética quantitativa. Centenas de novos genes que contribuem para a variação quantitativa em relação ao risco de doenças foram identificadas. Esses genes atualmente são o alvo para o desenvolvimento de novas terapias. Além dos seres humanos, os estudos de GWA avançaram a nossa compreensão sobre a herança de traços quantitativos em *Arabidopsis*, *Drosophila*, levedura e milho.

> CONCEITO-CHAVE O mapeamento de associação é um método para identificação de associações estatísticas entre marcadores moleculares e variação fenotípica em relação a traços complexos. O desequilíbrio de ligação em uma população entre o *locus* marcador e uma variante funcional

em um gene pode causar a associação. Se marcadores moleculares em todo o genoma estiverem disponíveis, um estudo de associação ampla do genoma (GWA) pode então ser realizado. Estudos de GWA em seres humanos possibilitaram que os geneticistas identificassem centenas de genes que contribuem para os riscos de desenvolvimento de muitas doenças comuns.

RESUMO

A genética quantitativa procura compreender a herança de traços complexos – os traços que são influenciados por uma mistura de fatores genéticos e ambientais e que não segregam em razões mendelianas simples. Os traços complexos podem ser traços categóricos, limiares, de contagem (merísticos) ou continuamente variáveis. Qualquer traço em relação ao qual não podemos inferir diretamente o genótipo a partir do fenótipo é um alvo para a análise genética quantitativa.

A arquitetura genética de um traço é a descrição total do número de genes que afetam o traço, das suas contribuições relativas para o fenótipo, da contribuição de fatores ambientais para o fenótipo e de uma compreensão sobre como os genes interagem uns com os outros e com os fatores ambientais. Para decifrar a arquitetura genética dos traços complexos, geneticistas quantitativos desenvolveram um modelo matemático simples, que decompõe os fenótipos dos indivíduos em diferenças que ocorrem em virtude de fatores genéticos (g) e aquelas que ocorrem em virtude de fatores ambientais (e).

As diferenças nos valores de traço entre os membros de uma população podem ser resumidas por uma medida estatística, denominada variância. A variância mede até que ponto os indivíduos se desviam da média da população. A variância de um traço pode ser separada em uma parte que ocorre em virtude de fatores genéticos (a variância genética) e uma parte que ocorre em virtude de fatores ambientais (a variância ambiental). Uma presunção-chave por trás da separação da variância do traço em componentes genéticos e ambientais é que os fatores genéticos e ambientais não estão correlacionados ou são independentes.

Até que ponto a variação em relação a um traço em uma população é explicada por fatores genéticos é medida pela herdabilidade no sentido amplo (H^2) do traço.

H^2 é a razão entre a variância genética e a variância fenotípica. A herdabilidade no sentido amplo expressa o grau até o qual as diferenças nos fenótipos entre os indivíduos em uma população são determinadas por diferenças nos seus genótipos. A medida de H^2 em seres humanos revelou que muitos traços apresentam influências genéticas, incluindo atributos físicos, funções mentais, características de personalidade, distúrbios psiquiátricos e até mesmo atitudes políticas.

Os genitores transmitem os genes, mas não os genótipos, aos seus descendentes. A cada geração, são criadas novas interações de dominância entre os alelos em um *locus*. Para incorporar esse fenômeno no modelo matemático em relação à variação quantitativa, o desvio genético (*g*) é decomposto nos desvios aditivo (*a*) e de dominância (*d*). Apenas o desvio aditivo é transmitido dos genitores para a descendência. O desvio aditivo representa a parte herdada do fenótipo no sentido restrito. A parte aditiva da variância em uma população é a parte herdada da variância. A herdabilidade no sentido restrito (h^2) é a razão da variância aditiva em relação à variância fenotípica. A herdabilidade no sentido restrito fornece uma medida da magnitude em que os fenótipos dos indivíduos são determinados pelos genes que eles herdam de seus genitores.

O conhecimento sobre a herdabilidade no sentido restrito de um traço é fundamental para compreender como um traço responderá ao cruzamento seletivo ou à força da seleção natural. Criadores de animais e cultivadores de plantas utilizam seu conhecimento sobre a herdabilidade no sentido restrito em relação aos traços de interesse para direcionar programas de melhoramento de plantas e animais. A herdabilidade no sentido restrito é utilizada para prever os fenótipos da descendência e para estimar o valor genético dos membros individuais da população reprodutora.

Os *loci* genéticos subjacentes à variação em traços complexos são conhecidos como *loci* de traços quantitativos ou QTL, abreviadamente. Existem dois métodos experimentais para a caracterização de QTL e a determinação de suas localizações no genoma. Primeiramente, o mapeamento de QTL procura por correlações estatísticas entre os genótipos em *loci* marcadores e os valores de traço em populações com heredogramas conhecidos, tais como uma população de

BC_1. O mapeamento de QTL proporciona estimativas sobre o número de genes que controlam um traço, se os alelos no QTL exibem aditividade ou dominância e se cada QTL apresenta um pequeno ou grande efeito sobre o traço. Em segundo lugar, o mapeamento de associação procura por correlações estatísticas entre os genótipos em *loci* marcadores e os valores de traço em populações de cruzamento aleatório. O mapeamento de associação pode possibilitar que pesquisadores identifiquem os genes que são a base do QTL. Estudos de associação ampla do genoma (GWA) utilizam marcadores que englobam o genoma inteiro.

A maior parte dos traços de importância na medicina, na agricultura e na biologia evolutiva demonstra herança complexa. Exemplos incluem o risco de doenças em seres humanos, a produção de soja, a produção de leite em vacas leiteiras e o espectro total de fenótipos que diferenciam todas as espécies de plantas, animais e microrganismos na terra. A análise genética quantitativa está à frente da compreensão da base genética desses traços críticos.

TERMOS-CHAVE

ação gênica
ação gênica aditiva
ação gênica dominante
ação gênica recessiva
amostra
arquitetura genética
associação ampla do genoma (GWA ou GWAS)
coeficiente de correlação
correlação
covariância
desvio
desvio padrão
diferencial de seleção (S)
distribuição normal
dominância parcial

efeito aditivo (A)
efeito da dominância (D)
gene candidato
genética quantitativa
genômica pessoal
herança complexa
herança simples
herdabilidade no sentido amplo (H^2)
herdabilidade no sentido restrito (h^2)
hipótese multifatorial
histograma de frequência
isogênico
linhagem congênica
linhagem ou cepa endocruzada
linhagem quase isogênica
loci de traço quantitativo (QTL)
mapeamento de associação
mapeamento de QTL
mapeamento fino
média
população
resposta à seleção (R)
traço categórico
traço complexo
traço contínuo
traço de limiar
traço merístico
traço quantitativo
valor produtivo
variância
variância aditiva
variância ambiental

variância da dominância
variância genética

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Em uma ninhada de frangos de corte, o peso médio é de 700 g e o desvio padrão é de 100 g. Presuma que os valores de traço sigam a distribuição normal.
a. Quantos dos frangos se espera que pesem mais de 700 g?
b. Quantos dos frangos se espera que pesem mais de 900 g?
c. Se H^2 é 1,0, qual é a variância genética em relação a essa população?

Solução

a. Tendo em vista que a distribuição normal é simétrica ao redor da média, 50% da população apresentará um valor de traço acima da média e os outros 50% apresentarão um valor de traço abaixo da média. Nesse caso, espera-se que 50 dos 100 frangos pesem mais de 700 g.
b. O valor de 900 g é 2 desvios padrões maior do que a média. Sob a distribuição normal, 95,5% da população encontram-se dentro de 2 desvios padrões da média e os 4,5% remanescentes estarão situados a mais de 2 desvios padrões da média. Desses 4,5%, metade (2,25%) será mais de 2 desvios padrões inferior à média e a outra metade (2,25%) será mais de 2 desvios padrões superior à média. Portanto, esperamos que aproximadamente 2,25% dos 100 frangos (ou aproximadamente 2 frangos) pesem mais de 900 g.
c. Quando H^2 é 1,0, toda a variação então é genética. Sabemos que o desvio padrão é 100 e a variância é o quadrado do desvio padrão.

$$\text{Variância} = \sigma^2$$

Portanto, a variação genética será $(100)^2 = 10.000\ g^2$.

Problema resolvido 2. Duas linhagens endocruzadas de feijões são intercruzadas. Na F_1, a variância no peso dos feijões é medida em 15 g^2. A F_1 é autofecundada;

na F_2, a variância no peso dos feijões é 61 g². Estime a herdabilidade ampla do peso dos feijões na população F_2 desse experimento.

Solução

Aqui, a chave é reconhecer que toda a variância na população F_1 deve ser ambiental, tendo em vista que todos os indivíduos apresentam o mesmo genótipo. Além disso, a variância na F_2 deve ser uma combinação de componentes ambientais e genéticos, tendo em vista que todos os genes que são heterozigotos na F_1 irão segregar na F_2 para proporcionar uma variedade de genótipos diferentes que estão relacionados com o peso dos feijões. Portanto, podemos estimar:

$$V_e = 15 \text{ g}^2$$
$$V_g + V_e = 61 \text{ g}^2$$

Portanto:

$$V_g = 61 - 15 = 46 \text{ g}^2$$

e a herdabilidade ampla é:

$$H^2 = \frac{46}{61} = 0{,}75 \ \ (75\%)$$

Problema resolvido 3. Em uma população experimental de *Tribolium* (besouros de farinha), o comprimento corporal demonstra uma distribuição contínua com média de 6 mm. Um grupo de machos e fêmeas com comprimento corporal médio de 9 mm é removido e intercruzado. O comprimento corporal de sua descendência apresenta média de 7,2 mm. A partir desses dados, calcule a herdabilidade no sentido restrito em relação ao comprimento corporal nessa população.

Solução

O diferencial da seleção (S) é 9 – 6 = 3 mm e a resposta à seleção (R) é 7,2 – 6 = 1,2 mm. Portanto, a herdabilidade no sentido restrito é:

$$h^2 = \frac{R}{S} = \frac{1{,}2}{3{,}0} = 0{,}4 \ \ (40\%)$$

Problema resolvido 4. Uma equipe de pesquisas relata que a herdabilidade no sentido amplo em relação à altura em seres humanos é 0,5 com base em um estudo de gêmeos idênticos criados em separado na Islândia. Outra equipe relata que a herdabilidade no sentido restrito em relação à altura humana é 0,8 com base em um estudo de correlação entre genitores e descendência nos EUA. O que parece ser inesperado a respeito desses resultados? Como os resultados inesperados podem ser explicados?

Solução

A herdabilidade no sentido amplo é a razão entre a variância genética total (V_g) e a variância fenotípica (V_X). A variância genética total inclui tanto a variância aditiva (V_a) quanto a de dominância (V_d):

$$h^2 = \frac{V_g}{V_X} = \frac{V_a + V_d}{V_X}$$

A herdabilidade no sentido restrito é a razão entre a variância aditiva (V_a) e a variância fenotípica (V_X).

$$H^2 = \frac{V_a}{V_X}$$

Portanto, todas as outras variáveis sendo iguais, H^2 deve ser maior ou igual a h^2. Ela será igual a h^2 quando V_d for 0,0. É inesperado que h^2 seja maior que H^2. Entretanto, as duas equipes de pesquisa estudaram diferentes populações – na Islândia e nos EUA. As estimativas de herdabilidade são aplicáveis apenas à população e ao ambiente no qual elas foram medidas. As estimativas realizadas em uma população podem ser diferentes daquelas realizadas em outra população, tendo em vista que duas populações podem segregar diferentes alelos em diversos genes e viver em ambientes diferentes.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. A Figura 19.9 demonstra as distribuições de traço antes e depois de um ciclo de seleção artificial. A variância do traço aparenta ter mudado como resultado da seleção?

2. A Figura 19.11 demonstra as distribuições esperadas em relação às três classes genotípicas se o *locus B* é um QTL que afeta o valor de traço.
a. Conforme desenhado, qual é a razão de dominância/aditiva (D/A)?
b. Como você redesenharia essa figura se o *locus B* não apresentasse efeito sobre o valor de traço?
c. Como as posições ao longo do eixo *x* das curvas em relação às diferentes classes genotípicas do *locus B* seriam alteradas se D/A = 1,0?

3. A Figura 19.16 demonstra os resultados de um experimento de mapeamento fino de QTL. Qual gene estaria implicado no controle do peso dos frutos se o peso médio dos frutos em relação a cada linhagem fosse como segue?

Linhagem	Peso do fruto (g)
1	181,4
2	169,3
3	170,7
4	171,2
5	171,4
6	182,2

7	180,6
8	180,7
9	181,8
10	169,3

4. A Figura 19.17 demonstra um conjunto de haplótipos. Suponha que esses sejam haplótipos em relação a um segmento cromossômico de 18 linhagens haploides de leveduras. Na extremidade direita da figura, S e D indicam se a linhagem sobrevive (S) ou morre (D) em alta temperatura (40°C). Com a utilização do teste do χ^2 (ver Capítulo 3) e da Tabela 3.1, o SNP1 ou o SNP6 demonstram evidências em relação a uma associação com o fenótipo de crescimento? Explique.

5. A Figura 19.18 A demonstra um gráfico de valores P (representados pelos pontos) ao longo dos cromossomos do genoma canino. Cada valor P é o resultado de um teste estatístico de associação entre um SNP e o tamanho corporal. Além do aglomerado de pequenos valores P próximos de *IGF1*, você observa algumas regiões cromossômicas com evidências de uma associação significativa entre um SNP e o tamanho corporal? Explique.

6. A Figura 19.19 demonstra gráficos de valores P (representados pelos pontos) ao longo dos cromossomos do genoma humano. Cada valor P é o resultado de um teste estatístico de associação entre um SNP e uma doença. Existe um aglomerado, ou espícula, de valores P estatisticamente significativos (*pontos verdes*) no gene *HLA-DRB1* em relação a duas doenças. Por que esse gene em particular pode contribuir para a suscetibilidade às doenças autoimunes artrite reumatoide e diabetes melito do tipo 1?

PROBLEMAS BÁSICOS

7. Diferencie entre a variação contínua e descontínua em uma população e forneça alguns exemplos de cada uma.

8. Quais são as premissas centrais da hipótese multifatorial?

9. A tabela abaixo demonstra uma distribuição do número de cerdas em uma população de *Drosophila*. Calcule a média, a variância e o desvio padrão em relação a esses dados.

Número de cerdas	Número de indivíduos
1	1
2	4
3	7
4	31
5	56
6	17
7	4

10. Suponha que o QI médio nos EUA seja de aproximadamente 100 e que o desvio padrão seja de 15 pontos. Pessoas com QI de 145 ou mais alto são consideradas “gênios” em algumas escalas de medição. Espera-se que qual porcentagem da população apresente um QI de 145 ou mais alto? Em um

país com 300 milhões de pessoas, quantos gênios se espera que existam?

11. Em uma amostra de mulheres adultas dos EUA, a altura média foi de 164,4 cm e o desvio padrão foi de 6,2 cm. As mulheres que estão a mais de 2 desvios padrões acima da média são consideradas muito altas e as mulheres que estão a mais de 2 desvios padrões abaixo da média são consideradas muito baixas. A altura em mulheres está distribuída normalmente.

a. Quais são as alturas das mulheres muito altas e muito baixas?

b. Em uma população de 10.000 mulheres, quantas se espera que sejam muito altas e quantas muito baixas?

12. Um cultivador de feijões está trabalhando com uma população na qual a quantidade média de vagens por planta é 50 e a variância é de 10 vagens². Sabe-se que a herdabilidade no sentido amplo é de 0,8. Tendo em vista essas informações, o cultivador pode ter certeza de que a população responderá à seleção em relação a um aumento no número de vagens por planta na próxima geração?

13. A tabela a seguir demonstra o número de leitões por ninhada em relação a um grupo de 60 porcas. Qual é a quantidade média de leitões por ninhada? Qual é a frequência relativa de ninhadas com no mínimo 12 leitões?

Número de ninhadas	Leitões/ninhada
1	6
3	7
7	8
12	9

18	10
20	11
17	12
14	13
6	14
2	15

14. Um granjeiro está trabalhando com uma população na qual o número médio de ovos postos por galinha em 1 mês é 28 e a variância é de 5 ovos². Sabe-se que a herdabilidade no sentido restrito é 0,8. Tendo em vista essas informações, o granjeiro pode esperar que a população responda à seleção em relação a um aumento no número de ovos por galinha na próxima geração?

a. Não, a aplicação da seleção é sempre arriscada e um granjeiro nunca sabe o que esperar.

b. Não, um granjeiro precisa conhecer a herdabilidade no sentido restrito para saber o que esperar.

c. Sim, tendo em vista que a herdabilidade no sentido restrito está próxima de 1 (0,8), pode-se então esperar que o cruzamento seletivo possa levar ao aumento da produção de ovos na próxima geração.

d. Sim, tendo em vista que a variância é superior a 0.

e. c e **d** estão corretas.

15. A herdabilidade no sentido restrito do número de ervilhas por vagem em uma população de ervilhas *sugar snap* é 0,5. A média da população é de

6,2 ervilhas por vagem. Um agricultor seleciona uma planta com 6,8 ervilhas por vagem e realiza o cruzamento com uma segunda planta que apresenta 8,0 ervilhas por vagem. Quais são os números esperados de ervilhas por vagem entre a descendência desse cruzamento?

16. O mapeamento de QTL e o mapeamento de GWA (de associação) são dois métodos diferentes utilizados para identificar genes que afetam traços complexos. Em relação a cada uma das declarações a seguir, escolha se ela é aplicável ao mapeamento de QTL, ao mapeamento de associação ou a ambos.

Declaração	**QTL**	**GWA**	**Ambos**
Esse método requer que o experimentador realize cruzamentos entre diferentes linhagens para produzir uma população de mapeamento			
Esse método pode varrer o genoma inteiro para encontrar QTL para um traço			
Esse método com frequência pode identificar os genes específicos que representam o QTL			
Esse método pode amostrar um grande número de indivíduos de uma população de cruzamento aleatório que apresentam variação em relação ao traço que está sendo estudado			

Esse método tipicamente testa dois alelos que diferem entre os dois genitores da população de mapeamento			

PROBLEMAS DESAFIADORES

17. Em um grande rebanho de bovinos, três diferentes características que demonstram distribuição contínua são medidas e são calculadas as variâncias na tabela a seguir:

	Características		
Variância	Comprimento da canela	Comprimento do pescoço	Conteúdo de gordura
Fenotípica	310,2	730,4	106,0
Ambiental	248,1	292,2	53,0
Genética aditiva	46,5	73,0	42,4
Genética de dominância	15,6	365,2	10,6

a. Calcule as herdabilidades no sentido amplo *e* restrito em relação a cada característica.

b. Na população de animais estudada, qual característica responderia

melhor à seleção? Por quê?

c. É realizado um projeto para diminuir o conteúdo médio de gordura no rebanho. O conteúdo médio de gordura atualmente é de 10,5%. Animais com média de conteúdo de gordura de 6,5% são intercruzados como genitores da próxima geração. Qual conteúdo médio de gordura pode ser esperado nos descendentes desses animais?

18. Em uma espécie de tentilhões (*Geospiza fortis*), a herdabilidade no sentido restrito da profundidade do bico foi estimada como sendo 0,79. A profundidade do bico está correlacionada com a capacidade dos tentilhões de ingerir sementes grandes. A profundidade média do bico para a população é de 9,6 mm. Um macho com uma profundidade de bico de 10,8 mm é cruzado com uma fêmea com uma profundidade de bico de 9,8 mm. Qual é o valor esperado em relação à profundidade do bico para a descendência desse cruzamento?

19. Duas linhagens endocruzadas de camundongos de laboratório são intercruzadas. Na F_1 (que apresenta genótipos idênticos em todos os *loci*), a variância no peso adulto é medida em 3 g^2. Os animais da F_1 são intercruzados para criar uma F_2 na qual a variância no peso adulto é de 16 g^2. Estime a herdabilidade ampla do peso adulto na população da F_2 desse experimento. (Os ambientes nos quais os animais da F_1 e da F_2 foram criados eram equivalentes.)

20. A tabela a seguir demonstra os pesos de 100 camundongos individuais da mesma linhagem endocruzada criados com dietas diferentes. Em relação a um camundongo que pesa 27 g, quanto de seu peso ocorre em virtude da sua genética e quanto da dieta específica que foi fornecida (ambiente)? (Com exceção da dieta, os camundongos foram criados em ambientes equivalentes.)

Número de camundongos	Peso (g)

5	21
13	22
18	23
21	24
22	25
16	26
5	27

21. A tabela a seguir contém as medidas do colesterol sérico total (mg/dℓ) em relação a 10 conjuntos de gêmeos monozigóticos que foram criados em separado. Calcule o seguinte: média geral, variância geral, covariância entre os gêmeos e herdabilidade no sentido amplo (H^2).

X′	X″
228	222
186	152
204	220

142	185
226	210
217	190
207	226
185	213
179	159
170	129

22. A tabela a seguir contém a altura em centímetros em relação a 10 conjuntos de mulheres gêmeas adultas. Calcule o coeficiente de correlação (r) entre as alturas das irmãs em relação aos pares de gêmeas.

Gêmea 1	Gêmea 2
158	163
156	150
172	173
156	154

160	163
159	153
170	174
177	174
165	168
172	165

---

23. A população A é composta por 100 galinhas que são totalmente isogênicas e que são criadas em um ambiente uniforme. O peso médio dos ovos que elas põem é de 52 g e a variância é de 3,5 g². A população B é composta por 100 galinhas geneticamente variáveis, que produzem ovos com um peso médio de 52 g e variância de 21,0 g². A população B é criada em um ambiente que é equivalente àquele da população A. Qual é a variância ambiental (V_e) em relação ao peso dos ovos? Qual é a variância genética na população B? Qual é a herdabilidade no sentido amplo na população B?

24. Plantas de milho em uma população apresentam em média 180 cm de altura. A herdabilidade no sentido restrito em relação à altura da planta nessa população é de 0,5. Um cultivador seleciona plantas que em média são 10 cm mais altas do que a média da população para produzir a próxima geração e o cultivador continua a aplicar esse nível de seleção durante oito gerações. Qual será a altura média das plantas após oito gerações de seleção? Presuma que h^2 permanece 0,5 e que V_e não é alterada ao longo do período do experimento.

25. Em uma população de *Drosophila melanogaster* criada em laboratório, o

comprimento médio das asas é de 0,55 mm e a variação é de 0,35 a 0,65. Um geneticista seleciona uma fêmea com asas que têm 0,42 mm de comprimento e a cruza com um macho que tem asas com 0,56 mm de comprimento.

a. Qual é o comprimento esperado das asas de sua descendência se o comprimento das asas apresenta uma herdabilidade no sentido restrito de 1,0?

b. Qual é o comprimento esperado das asas de sua descendência se o comprimento das asas apresenta uma herdabilidade no sentido restrito de 0,0?

26. Diferentes espécies de grilos apresentam cantos distintos e eles utilizam esses cantos para o reconhecimento dos parceiros. Pesquisadores cruzaram duas espécies de grilos havaianos (*Laupala paranigra* e *L. kohalensis*), cujos cantos são distinguidos pela frequência de pulso (o número de pulsos/s; Shaw *et al.*, *Molecular Ecology* 16, 2007, 2879-2892). Em seguida, eles mapearam QTL na população F_2 derivada desse cruzamento. Foram detectados seis QTL autossômicos. Os valores de traço médio (pulsos/s) nas três classes genotípicas da F_2 em relação a cada QTL estão demonstrados na tabela a seguir, em que P indica o alelo de *L. paranigra* e K indica o alelo de *L. kohalensis*.

QTL	P/P	P/K	K/K
1	1,54	1,89	2,10
2	1,75	1,87	1,94
3	1,72	1,88	1,92

4	1,70	1,82	2,02
5	1,67	1,80	2,13
6	1,57	1,88	2,19

a. Calcule os efeitos aditivos (A) e de dominância (D) e a razão de D/A em relação a cada um dos seis QTL.

b. Qual desses QTL demonstra a maior quantidade de dominância?

c. Qual desses apresenta o maior efeito aditivo?

d. A frequência de pulso média em relação ao *L. kohalensis* é de 3,72 e é de 0,71 em relação ao *L. paranigra*. Todos os seis QTL atuam no sentido esperado, com o alelo de *L. kohalensis* conferindo uma frequência de pulso mais alta do que o alelo de *L. paranigra*?

27. A questão 26 faz referência ao QTL nos autossomos de grilos. Em relação aos cromossomos sexuais, grilos-fêmeas são XX e grilos-machos são XO, com a apresentação de apenas um cromossomo X, mas nenhum cromossomo Y. O QTL em relação à frequência de pulso pode ser mapeado em cromossomos X de grilos? Se o canto é realizado apenas por grilos-machos, os efeitos de dominância do QTL sobre o X podem ser estimados?

28. Estudos de GWA revelam correlações estatísticas entre os genótipos em *loci* marcadores em genes e traços complexos. Os estudos de GWA comprovam que a variação alélica em um gene realmente causa a variação no traço? Caso negativo, quais experimentos poderiam comprovar que as variantes alélicas em um gene em uma população são responsáveis pela variação em um traço?

29. O gene do *albinismo ocular-2* (*OCA2*) e o gene do *receptor 1 de melanocortina* (*MC1R*) estão, ambos, envolvidos no metabolismo da melanina em células cutâneas em seres humanos. Para testar se uma variação nesses genes contribui para a sensibilidade ao sol e o risco

correlato de ser afetado pelo câncer de pele, você realiza análises de associação. Solicitamos a uma amostra de 1.000 pessoas da Islândia que classificassem a si próprias como bronzeadas ou com queimaduras (nenhum bronzeamento) quando expostas ao sol. Os indivíduos também foram genotipados em relação a um SNP em cada gene (rs7495174 e rs1805007). A tabela demonstra o número de indivíduos em cada classe.

	OCA2 (rs7495174)			*MC1R* (rs1805007)		
	A/A	A/G	G/G	C/C	C/T	T/T
Queimadura	245	56	1	192	89	21
Bronzeamento	555	134	9	448	231	19

a. Quais são as frequências dos fenótipos de bronzeamento e queimadura na Islândia?

b. Quais são as frequências alélicas em cada *locus* (SNP)?

c. Com a utilização do teste do χ^2 (ver Capítulo 3) e da Tabela 3.1, teste a hipótese nula de que não existe associação entre esses SNP e a pele sensível ao sol. Algum SNP demonstra evidência em relação a uma associação?

d. Se você observar evidências de uma associação entre o gene e o traço, qual é o modo de ação gênica?

e. Se o valor P for superior a 0,05, isso comprova que o gene não contribui para a variação em relação à sensibilidade ao sol? Por quê?

CAPÍTULO 20

Evolução de Genes e Traços

A teoria da evolução pela seleção natural foi desenvolvida independentemente por dois intrépidos naturalistas britânicos, Charles Darwin (1809-1882) e Alfred Russel Wallace (1823-1913), no período de suas respectivas longas viagens. (*Esquerda*, Graphic Arts/Corbis; *direita*, Hulton Archive/Getty Images.)

TÓPICOS

20.1 Evolução por seleção natural
20.2 Seleção natural em ação | Caso exemplar

20.3 Evolução molecular | Teoria neutra
20.4 Seleção cumulativa e vias de múltiplas etapas até a alteração funcional
20.5 Evolução morfológica
20.6 Origem de novos genes e funções proteicas

RESULTADOS DE APRENDIZAGEM

Após ler este capítulo, você será capaz de:

- Identificar e explicar os ingredientes essenciais da evolução por seleção natural
- Descrever exemplos de traços e genes que foram desenvolvidos por meio da seleção natural
- Contrastar a evolução molecular neutra com os processos adaptativos
- Diferenciar as assinaturas da seleção positiva e da seleção purificadora nas sequências de DNA e proteínas
- Contrastar análises experimentais e estatísticas da seleção cumulativa em proteínas
- Formular uma justificativa para o papel crítico das sequências reguladoras na evolução de traços morfológicos
- Avaliar o papel da duplicação de genes na origem de novas funções proteicas.

Charles Darwin (1809-1882) chegou às ilhas Galápagos em 1835, no quarto ano do que deveria ter sido uma viagem de dois anos. Pode-se pensar que essas ilhas, agora intrinsecamente ligadas ao nome de Darwin, fossem o paraíso para o jovem naturalista. Longe disso. Darwin achou as ilhas terrivelmente quentes, com suas rochas magmáticas negras queimando sob o sol quente. Em seu diário, ele observou que “as árvores raquíticas demonstram poucos sinais de vida (...) as plantas também têm um odor desagradável (...) As rochas magmáticas negras na praia são frequentadas por grandes (60 a 90 cm) lagartos, os mais desajeitados e repugnantes (...) Eles certamente dominam o

terreno que habitam".[1] Além dos lagartos e das tartarugas, a vida animal nas ilhas era escassa e inexpressiva. Ele mal podia esperar para ir embora. O explorador de 26 anos de idade não sabia que suas 5 semanas em Galápagos inspirariam uma série de ideias radicais que, aproximadamente 24 anos depois, com a publicação de seu *A Origem das Espécies* (1859), alterariam a nossa percepção sobre o mundo e o nosso lugar nele.

Diversos meses após deixar as ilhas, na última parte da viagem de volta à Inglaterra, Darwin teve seu primeiro lampejo de percepção. Ele havia começado a organizar suas abundantes anotações de campo de seus quase 5 anos de exploração e coleta. Seu plano era que de volta à Inglaterra especialistas liderassem o estudo de suas coleções de fósseis, plantas, animais e rochas. Voltando-se para as suas observações sobre as aves de Galápagos, ele relembrou que havia observado tipos ligeiramente diferentes de tordos em três ilhas diferentes. Agora, havia um quebra-cabeça. A visão prevalecente sobre a origem das espécies em 1835, sustentada pela maior parte dos professores de Darwin e por uma grande parte da comunidade científica, era de que as espécies haviam sido especialmente criadas por Deus na sua forma atual, inalterável, e inseridas no *habitat* ao qual elas se adaptavam melhor. Por que, então, haveria aves ligeiramente diferentes em ilhas tão semelhantes? Darwin anotou em seu caderno de ornitologia:

> Quando vejo essas ilhas, uma à vista das outras e possuídas por nada além de um escasso estoque de animais, habitadas por essas aves, que diferem ligeiramente na estrutura que preenche o mesmo espaço na Natureza, devo suspeitar de que sejam apenas variedades.
> (…) Se houver um mínimo fundamento nessas anotações, bem valerá a pena examinar a zoologia dos Arquipélagos, *tendo em vista que tais fatos abalariam a estabilidade das espécies* [grifo nosso].[2]

A percepção de Darwin foi de que as espécies podem ser alteradas. Isso não era o que ele havia aprendido em Cambridge University. Era uma heresia. Embora Darwin tivesse decidido manter para si próprio tais perigosos pensamentos, ele foi absorvido pela ideia. Após chegar em casa na Inglaterra, ele preencheu uma série de cadernos com pensamentos a respeito da alteração das espécies. Dentro de um ano, ele havia se convencido de que as espécies surgem naturalmente a partir de espécies preexistentes, tão naturalmente quanto as crianças nascem dos

pais e os pais, dos avós. Ele então ponderou *como* as espécies são alteradas e se adaptam às circunstâncias em particular. Em 1838, apenas dois anos após a conclusão de sua viagem e antes que ele completasse 30 anos, ele concebeu a sua resposta – **seleção natural**. Nesse processo competitivo, indivíduos que apresentam alguma vantagem relativa sobre outros vivem por mais tempo e produzem mais descendência, a qual, por sua vez, herda a vantagem.

Darwin sabia que, para convencer outras pessoas de suas ideias – a descendência das espécies a partir de ancestrais e a seleção natural –, ele precisaria de mais evidências. Passou as duas próximas décadas organizando todos os fatos que ele conseguia a partir de registros de botânica, zoologia, embriologia e fósseis.

Ele recebeu informações cruciais de especialistas que ajudaram a separar e caracterizar as suas coleções. O ornitólogo John Gould apontou para Darwin que o que o jovem naturalista acreditava serem melros, pardais e tentilhões de Galápagos eram, de fato, 12 (atualmente reconhecidas como 13) novas e distintas espécies de tentilhões terrestres (Figura 20.1). As espécies de Galápagos, embora claramente consideradas tentilhões, exibem uma imensa variação no comportamento alimentar e no formato do bico, que corresponde às suas fontes de alimentos. Por exemplo, o tentilhão-arborícola-vegetariano utiliza seu bico forte para comer frutos e folhas, o tentilhão insetívoro apresenta um bico com uma ponta cortante para comer grandes insetos e, o mais extraordinário de tudo, o tentilhão-pica-pau agarra um ramo no seu bico e o utiliza para obter insetos ao examinar os buracos nas árvores.

Essa diversidade de espécies, Darwin deduziu, deve ter surgido a partir de uma população original de tentilhões que chegou a Galápagos a partir do continente da América do Sul e povoou as ilhas. Os descendentes dos colonizadores originais se difundiram até as diferentes ilhas e formaram populações locais, que divergiram entre si e que finalmente formaram diferentes espécies.

Os tentilhões ilustram o processo de **adaptação**, no qual as características de uma espécie se modificam para a adaptação ao ambiente em que vivem. Darwin forneceu um nível de explicação para o processo, a seleção natural, mas não

conseguiu explicar como os traços variavam ou como eles foram alterados com o tempo, tendo em vista que ele não compreendia os mecanismos de herança. A compreensão da base genética da adaptação é um dos objetivos a longo prazo da biologia evolutiva.

Um primeiro passo em direção a esse objetivo foi realizado quando o trabalho de Mendel que apontava para a existência dos genes foi redescoberto duas décadas após a morte de Darwin. Outra chave surgiu herança meio século mais tarde, quando a base molecular da hereditabilidade e o código genético foram decifrados. Há muitas décadas, biólogos têm conhecimento de que as espécies e os traços se desenvolvem por meio de alterações na sequência de DNA. Entretanto, a elucidação das alterações específicas na sequência de DNA que levam à evolução fisiológica ou morfológica impôs desafios técnicos consideráveis. Avanços em genética molecular, genética do desenvolvimento e genômica comparativa estão agora revelando os diversos mecanismos subjacentes à evolução gênica, dos traços e da diversidade dos organismos.

Uma diversidade de espécies pode resultar da adaptação

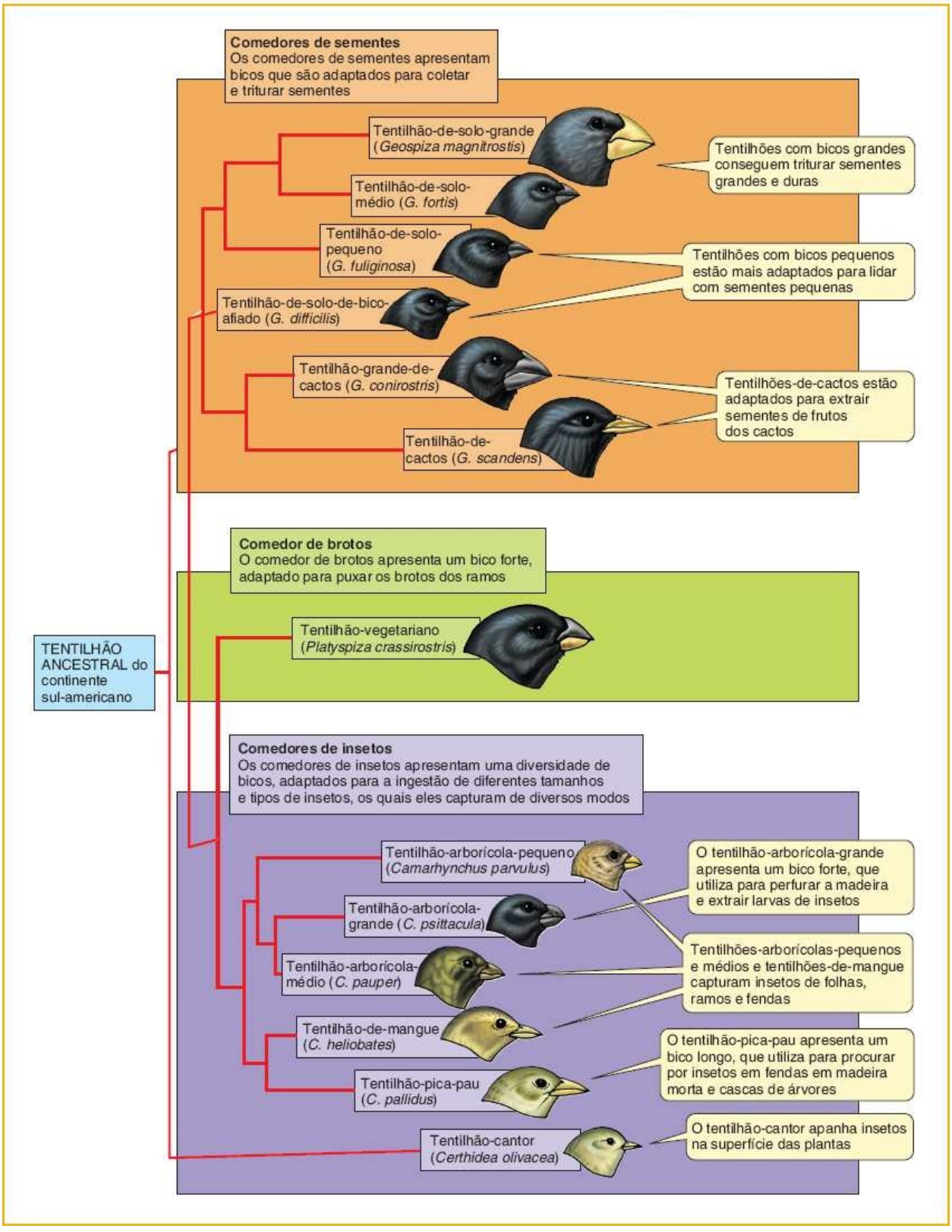


FIGURA 20.1 As 13 espécies de tentilhões encontradas nas ilhas Galápagos.

O estudo da evolução é uma disciplina muito ampla e em expansão. Como tal, não tentaremos uma visão geral abrangente de todas as facetas da análise evolutiva. Em vez disso, neste capítulo, examinaremos os mecanismos genéticos moleculares que levam à variação e à evolução dos traços, bem como à adaptação dos organismos aos seus ambientes. Primeiramente, examinaremos o processo evolutivo em geral e, em seguida, focaremos em exemplos específicos em relação aos quais foram apontadas as bases genéticas e moleculares das diferenças fenotípicas entre as populações ou as espécies. Todos os exemplos enfocarão a evolução de traços relativamente simples, controlados por um único gene. Esses exemplos relativamente simples são suficientes para ilustrar o processo fundamental de evolução no nível do DNA e a diversidade de maneiras pelas quais a evolução gênica afeta o ganho, a perda e a modificação dos traços.

20.1 Evolução por seleção natural

A teoria da evolução moderna está tão completamente identificada com o nome de Darwin que muitas pessoas pensam que o próprio Darwin tenha proposto pela primeira vez o conceito de que os organismos evoluíam, mas esse não é o caso. A ideia de que a vida foi alterada ao longo do tempo estava circulando nas comunidades científicas havia muitas décadas antes da histórica viagem de Darwin. A grande questão era: "*Como* a vida era alterada?" Para alguns, a explicação era uma série de criações especiais de Deus. Para outros, tais como Jean-Baptiste Lamarck (1744-1829), a alteração era causada pelo ambiente que atuava diretamente sobre o organismo, e aquelas alterações adquiridas no período de vida de um organismo eram transmitidas para a sua descendência.

O que Darwin proporcionou foi uma explicação detalhada sobre o mecanismo do processo evolutivo que incorporava corretamente o papel da herança. A teoria da evolução por seleção natural de Darwin tem início com a variação que existe entre os organismos de uma espécie. Indivíduos de uma geração são qualitativamente diferentes entre si. A evolução da espécie como um todo resulta

do fato de que os diversos tipos diferem em suas taxas de sobrevivência e reprodução. Tipos mais bem-adaptados deixam mais descendência e, assim, as frequências relativas dos tipos são alteradas ao longo do tempo. Portanto, os três ingredientes críticos da alteração evolutiva que Darwin apresentou foram a variação, a seleção e o tempo:

> Pode, então, ser considerado improvável (...) que variações de algum modo úteis para cada ser na grande e complexa batalha da vida, por vezes possam ocorrer no período de milhares de gerações? (...) Podemos duvidar (relembrando que nascem muito mais indivíduos do que possivelmente conseguem sobreviver) de que indivíduos que apresentam qualquer vantagem sobre outros, ainda que pequena, apresentam maior chance de sobreviver e procriar a sua espécie? Por outro lado, podemos ter a certeza de que qualquer variação no mínimo prejudicial será rigidamente destruída. A essa preservação de variações favoráveis e rejeição de variações prejudiciais, eu dou o nome de Seleção Natural. (*A Origem das Espécies*, Capítulo IV)[3]

Os textos e as ideias de Darwin são bem-conhecidos, e é justificável que sejam, mas é muito importante observar que ele não estava sozinho ao chegar até esse conceito de seleção natural. Alfred Russel Wallace (1823-1913), um colega inglês que explorou a floresta amazônica e o arquipélago Malaio durante 12 anos, chegou a uma conclusão muito semelhante em um artigo que publicou em conjunto com um trecho de Darwin em 1858:

> A vida dos animais selvagens é uma luta pela existência (...) Talvez todas as variações da forma típica de uma espécie devam apresentar algum efeito definitivo, mas leve, sobre os hábitos ou as capacidades dos indivíduos (...) Também está evidente que a maior parte das alterações afeta, seja de modo favorável ou adverso, a possibilidade de uma existência prolongada (...) Se, por outro lado, qualquer espécie precisar produzir uma variedade que apresente capacidade ligeiramente aumentada para a preservação da existência, aquela variedade inevitavelmente deve, no tempo, adquirir uma superioridade numérica.[4]

Embora a denominação de Darwin atualmente tenda a ser exclusivamente ligada à evolução por seleção natural, na sua época, a teoria foi reconhecida como teoria de Darwin-Wallace. Talvez a atual percepção ocorra, pelo menos em parte, em virtude do próprio Wallace, que sempre fazia deferências a Darwin e se referia à teoria da evolução emergente como “darwinismo”.

> CONCEITO-CHAVE Darwin e Wallace propuseram uma nova explicação

> para justificar o fenômeno da evolução. Eles compreenderam que a população de uma determinada espécie em uma determinada ocasião inclui indivíduos com características variáveis. Perceberam que a população de gerações sucessivas conterá uma frequência mais alta daqueles tipos que sobrevivem e se reproduzem com mais sucesso sob as condições ambientais existentes. Portanto, as frequências dos diversos tipos dentro das espécies serão alteradas ao longo do tempo.

Existe uma óbvia semelhança entre o processo evolutivo, conforme Darwin e Wallace o descreveram, e o processo por meio do qual o agricultor ou pecuarista melhora um estoque doméstico. O agricultor seleciona as plantas de mais alta produção da população atual e as utiliza como genitoras da próxima geração. Se as características que causam a mais alta produção são hereditárias, a próxima geração então deverá ter produção mais alta. Não foi ao acaso que Darwin escolheu o termo *seleção natural* para descrever o seu modelo de evolução por meio de diferenças nas taxas de reprodução demonstradas por diferentes variantes na população. Como um modelo para esse processo evolutivo na natureza, ele tinha em mente a seleção que os cultivadores e criadores exercem sobre sucessivas gerações de plantas e animais domésticos.

Podemos resumir a teoria da evolução pela seleção natural em três princípios:

1. *Princípio da variação.* Entre os indivíduos de qualquer população, existe variação na morfologia, na fisiologia e no comportamento.
2. *Princípio da hereditariedade.* A descendência se assemelha aos seus genitores mais do que se assemelha aos indivíduos não relacionados.
3. *Princípio da seleção.* Alguns tipos obtêm mais sucesso na sobrevivência e na reprodução do que outros tipos em um determinado ambiente.

Um processo seletivo pode produzir alterações na composição da população apenas se houver algumas variações que possam ser selecionadas. Se todos os indivíduos forem idênticos, nenhuma diferença nas taxas reprodutivas dos indivíduos, não importa quanto extrema seja, alterará a composição da população.

Além disso, a variação deve ser em parte hereditária para que essas diferenças nas taxas reprodutivas alterem a composição genética da população. Se animais grandes de uma população apresentarem mais descendência do que os animais pequenos, mas a sua descendência em média não for maior do que aquela dos animais pequenos, não haverá então alteração na composição da população de uma geração até outra. Finalmente, se todos os tipos variantes deixarem, em média, o mesmo número de descendentes, podemos então esperar que a população permaneça inalterada.

> CONCEITO-CHAVE Os princípios da variação, da hereditariedade e da seleção devem ser todos aplicáveis para que ocorra evolução por meio de um mecanismo de variação.

A variação hereditária proporciona a matéria-prima para alterações sucessivas em uma espécie e para a multiplicação de novas espécies. Os mecanismos básicos dessas alterações (conforme discutido no Capítulo 18) são a origem de novas variações genéticas por mutação, da alteração na frequência de alelos em populações por meio de processos seletivos e aleatórios, da divergência de diferentes populações em virtude da diferença das forças seletivas ou em virtude da deriva aleatória e da redução da variação entre populações por migração (Figura 20.2). A partir desses mecanismos básicos, pode ser derivado um conjunto de princípios que regulam as alterações na composição genética das populações. A aplicação desses princípios da genética de populações proporciona uma teoria genética da evolução.

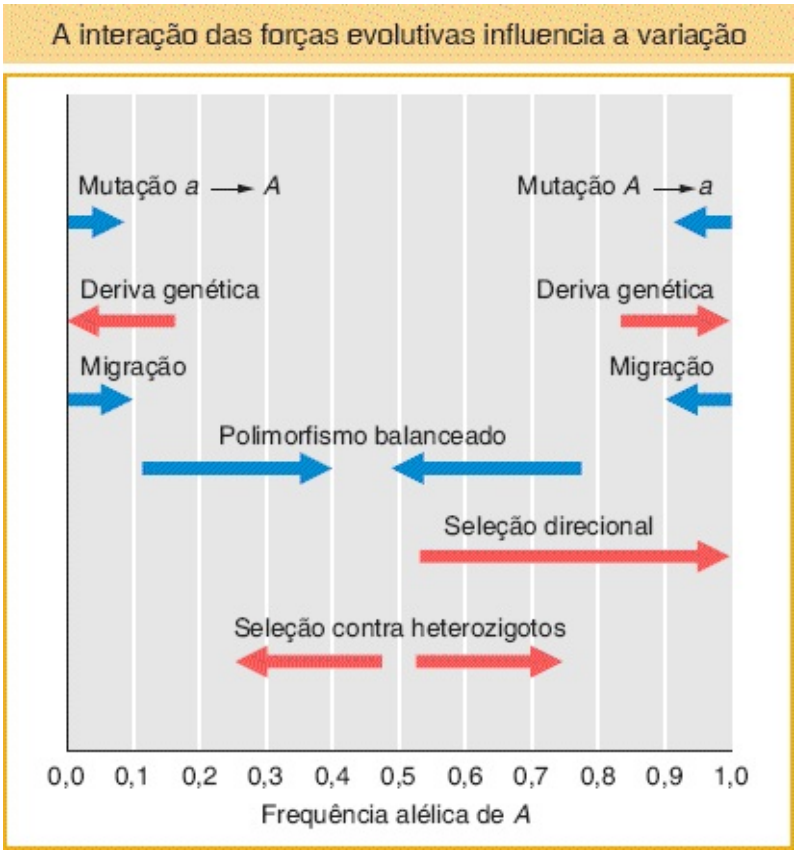


FIGURA 20.2 Os efeitos de diversas forças evolutivas sobre a frequência alélica. As *setas azuis* demonstram uma tendência em direção ao aumento da variação na população; as *setas vermelhas*, diminuição da variação.

CONCEITO-CHAVE A evolução, a alteração em populações ou espécies ao longo do tempo, é a conversão da variação hereditária entre indivíduos nas populações em diferenças hereditárias entre populações no tempo e no espaço por meio de mecanismos da genética de populações.

20.2 Seleção natural em ação | Caso exemplar

Durante quase um século após a publicação de *A Origem das Espécies*, não houve um exemplo de seleção natural que não tenha sido totalmente elucidado; ou seja, onde o agente da seleção natural era conhecido, o efeito sobre os diferentes genótipos pôde ser medido, a base genética e molecular da variação foi

identificada, e o papel fisiológico do gene ou da proteína envolvida foi bem-compreendido.

O primeiro exemplo “integrado” de seleção natural sobre uma variante molecular foi elucidado na década de 1950, antes mesmo da elucidação do código genético. Notavelmente, esse trabalho pioneiro revelou que a seleção natural opera em seres humanos. Ele ainda permanece como um dos exemplos mais detalhados e importantes de evolução por seleção natural em qualquer espécie.

A história teve início quando Tony Allison, um estudante de medicina de Oxford nascido no Quênia, fez um estudo de campo dos tipos sanguíneos entre tribos quenianas. Um dos exames de sangue que ele realizou foi em relação às células falciformes, eritrócitos que adquirem um formato de foice com a exposição ao agente redutor metabissulfito de sódio ou após ficarem em repouso por alguns dias (Figura 20.3). As células deformadas são um marco da anemia falciforme, uma doença descrita pela primeira vez em 1910. Essas células causam complicações patológicas por meio da oclusão dos vasos sanguíneos e levam à mortalidade precoce.

Em 1949, o mesmo ano em que Allison realizava seu trabalho de campo, o grupo de pesquisa de Linus Pauling demonstrou que pacientes com anemia falciforme apresentavam proteína hemoglobina com uma carga anormal (hemoglobina S, ou HbS) em seu sangue, em comparação à hemoglobina de indivíduos não afetados (hemoglobina A, ou HbA). Essa foi a primeira demonstração de uma anormalidade molecular ligada a uma doença complexa. Naquela época, em geral compreendia-se que portadores de células falciformes eram heterozigotos e, portanto, apresentavam uma mistura de HbA e HbS (indicados como *AS*), enquanto os indivíduos afetados eram homozigotos em relação ao alelo Hb^S (indicados como *SS*).

Allison coletou amostras de sangue de membros das tribos Kikuyo, Masai, Luo e outras tribos em diversas regiões do Quênia. Embora ele não tenha observado associação particularmente surpreendente entre os tipos sanguíneos AB0 ou MN entre as tribos, ele mediu frequências notavelmente diferentes de Hb^S. Nas tribos que viviam no Quênia central árido ou nas montanhas, a frequência de Hb^S era

inferior a 1%; entretanto, nas tribos que viviam na costa ou próximo do lago Victoria, a frequência de Hb^S normalmente excedia os 10% e se aproximava de 40% em alguns locais (Tabela 20.1).

As frequências alélicas eram surpreendentes por dois motivos. Primeiramente, tendo em vista que a anemia falciforme normalmente era letal, por que as frequências do alelo Hb^S eram tão altas? E, em segundo lugar, tendo em vista as distâncias relativamente curtas entre as regiões, por que a frequência de Hb^S era tão alta em alguns locais e em outros não?

A familiaridade de Allison com o terreno, as tribos e as doenças tropicais do Quênia o levou à explicação crucial. Allison percebeu que o alelo Hb^S estava a uma frequência alta em regiões úmidas e baixas, com níveis muito altos de malária, e quase ausente em altitudes elevadas, tais como ao redor de Nairóbi. Carregado por mosquitos, o parasita intracelular *Plasmodium falciparum*, que causa a malária, multiplica-se dentro dos eritrócitos (Figura 20.4). Os mosquitos e a doença são prevalentes por toda a África Subsaariana em regiões baixas e úmidas, próximas a coleções de água, nas quais os mosquitos se reproduzem. Allison conjeturou que o alelo Hb^S poderia, por meio da alteração dos eritrócitos, conferir algum grau de resistência à infecção por malária.

Tabela 20.1 Frequência de Hb^S em algumas tribos quenianas.

Tribo	Afinidade ética	Distrito/região	% Hb^S
Luo	Nilótica	Kisumu (lago Victoria)	25,7
Suba	Bantu	Ilha Rusingo	27,7
Kikuy	Bantu	Nairóbi	0,4

FIGURA 20.3 Uma micrografia eletrônica colorida demonstrando células falciformes entre eritrócitos normais. (Eye of Science/Science Source.)

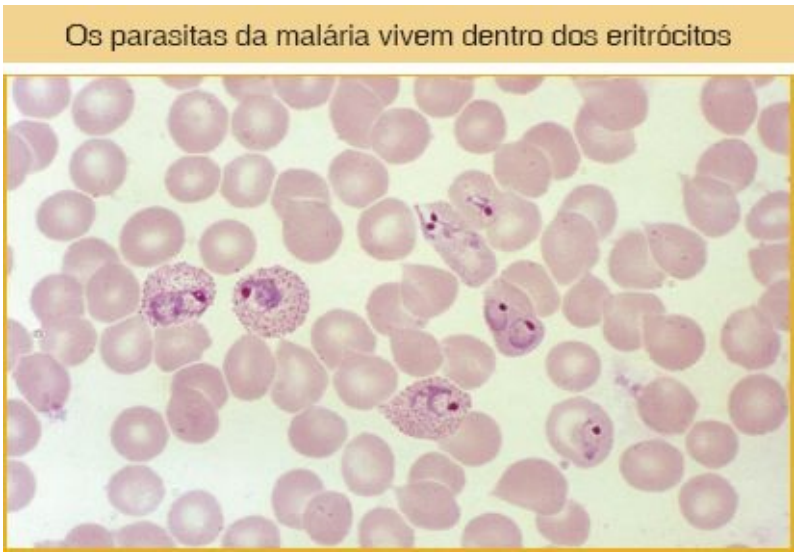


FIGURA 20.4 Esfregaço sanguíneo de um indivíduo infectado por parasitas da malária. Uma amostra eritrocitária foi tratada com corante Giemsa para revelar os parasitas dentro das células (*pontos vermelhos*). (CDC/Dra. Mae Melvin.)

Vantagem seletiva de Hb^S

Com a finalidade de testar esta ideia, Allison realizou uma pesquisa muito maior da frequência de *Hb*^{S} por toda a África Oriental, incluindo Uganda, Tanzânia e

Quênia. Ele examinou aproximadamente 5.000 indivíduos, que representavam mais de 30 tribos diferentes. Novamente, ele observou frequências de Hb^S de até 40% em áreas com malária e frequências de até 0% onde a malária estava ausente.

A ligação sugeria que o alelo Hb^S poderia afetar os níveis do parasita e, assim, Allison realizou um estudo sobre o nível de parasitas no sangue de crianças heterozigotas *AS versus* crianças do tipo selvagem *AA*. Em um estudo de quase 300 crianças, ele observou que a incidência de parasitas da malária de fato era inferior em crianças *AS* (27,9%) do que em crianças *AA* (45,7%) e que a densidade de parasitas também era mais baixa em crianças *AS*. Os resultados indicaram que crianças *AS* apresentavam mais baixas incidência e gravidade de infecções por malária e que, portanto, apresentavam uma vantagem seletiva nas áreas em que a malária era prevalente.

A vantagem dos heterozigotos *AS* era especialmente surpreendente à luz da doença sofrida pelos homozigotos *SS*. Allison observou:

> A proporção de indivíduos com células falciformes em qualquer população, então, será o resultado de um equilíbrio entre dois fatores: a gravidade da malária, que tenderá a aumentar a frequência do gene, e a taxa de eliminação dos genes falciformes em indivíduos que morrem por anemia falciforme (…) Geneticamente falando, esse é um *polimorfismo balanceado* [grifo nosso], no qual o heterozigoto apresenta uma vantagem sobre qualquer homozigoto.[5]

Em outras palavras, a mutação falciforme estava sob *seleção balanceada* (ver Capítulo 18) nas áreas em que a malária estava presente. A seleção positiva que opera sobre indivíduos *AS* é equilibrada pela seleção natural que opera contra os indivíduos *AA* suscetíveis à malária e os indivíduos *SS* que podem sucumbir à anemia falciforme.

Quanta vantagem os indivíduos *AS* vivenciam? Isso pode ser calculado ao medir a frequência do alelo Hb^S nas populações e examinar como essas frequências diferem das frequências esperadas sob as presunções da equação de Hardy-Weinberg (ver Capítulo 18). Uma análise em grande escala de 12.387 africanos orientais revelou uma frequência do alelo Hb^S (q) de 0,123. As frequências calculadas a partir da equação de Hardy-Weinberg são mais baixas em relação aos fenótipos homozigotos e mais altas em relação ao fenótipo

heterozigoto (Tabela 20.2). Se presumirmos que o heterozigoto *AS* apresenta adaptabilidade de 1,0, o valor adaptativo relativo dos outros genótipos então pode ser estimado a partir dessas diferenças. O valor adaptativo relativo do genótipo heterozigoto *AS* é 1,0/0,88 = 1,136, que corresponde a uma vantagem seletiva de aproximadamente 14%.

Essa vantagem seletiva foi bem-documentada por estudos da sobrevida a longo prazo de crianças *AA*, *AS* e *SS* no Quênia. Esses estudos observaram que os indivíduos *AS* apresentam uma vantagem de sobrevida pronunciada sobre indivíduos *AA* e *SS* nos primeiros anos de vida (Figura 20.5).

> CONCEITO-CHAVE O alelo da hemoglobina falciforme, Hb^S, está sob seleção balanceada em zonas com malária e transmite uma grande vantagem de sobrevida em heterozigotos ao longo dos primeiros anos de vida.

Origens moleculares de Hb^S

Após a descoberta de Allison, houve um entusiasmado interesse em determinar a base molecular da(s) diferença(s) entre Hb^S e Hb^A. O sequenciamento da proteína determinou que Hb^S difere de Hb^A em apenas um aminoácido, uma valina no lugar de um resíduo de ácido glutâmico. Essa alteração de um único aminoácido altera a carga das hemoglobinas e causa a sua agregação em estruturas longas em formato de bastão dentro dos eritrócitos. Após o código genético ter sido decifrado e métodos para o sequenciamento do DNA terem sido desenvolvidos, determinou-se que Hb^S é causada por uma única mutação de ponto (CTC → CAC) no códon do ácido glutâmico, que codifica o sexto aminoácido da subunidade β-globina dentro da proteína hemoglobina.

Tabela 20.2 Vantagem em termos de adaptabilidade de heterozigotos falciformes.

Genótipo	Frequência fenotípica observada	Frequência fenotípica esperada	Razão de observada/esperada	*W* (adaptabi relativa)

SS	29	187,4	0,155	0,155/1,12 =
AS	2.993	2.672,4	1,12	1,12/1,12 =
AA	9.365	9.527,2	0,983	0,983/1,12 =
Total	12.387	12.387		

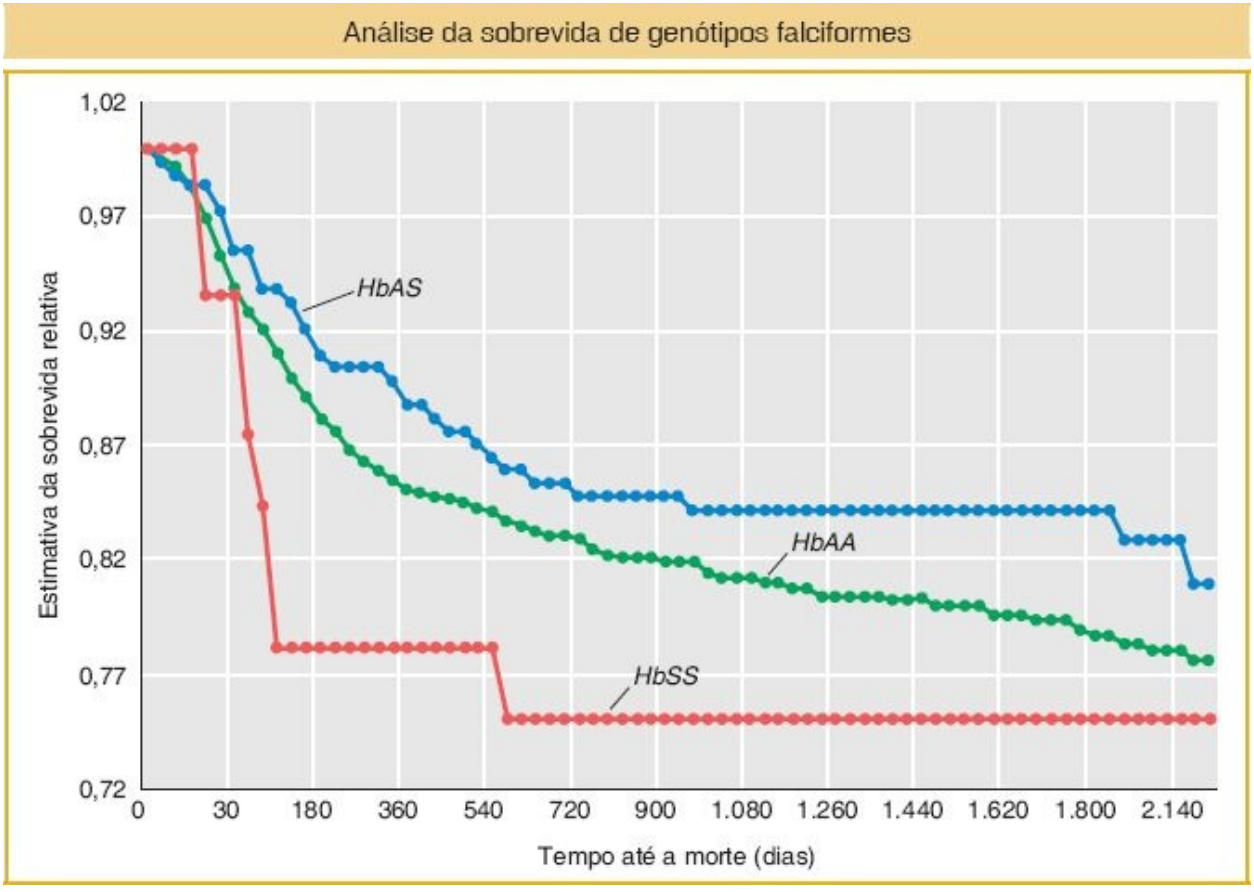


FIGURA 20.5 A sobrevida relativa de aproximadamente 1.000 crianças de Kisumu está inserida em gráfico desde o nascimento até os primeiros anos de vida. Heterozigotos falciformes apresentaram uma vantagem significativa na sobrevida geral dos 2 aos 16 meses de idade. (Dados de M. Aidoo *et al.*, *The Lancet* 359, 2002, 1311-1312.)

Curiosamente, Allison também observou uma alta incidência de Hb^S fora da África, incluindo Itália, Grécia e Índia. Outros marcadores de tipo sanguíneo não indicaram fortes relações genéticas entre essas populações. Em vez disso, Allison observou que essas também eram áreas com alta incidência de malária. A correlação entre a frequência de Hb^S e a incidência de malária era mantida não apenas na África Oriental, mas no continente africano, no sul da Europa e no subcontinente indiano. Allison compôs mapas demonstrando essas surpreendentes correlações (Figura 20.6) e inferiu que os alelos Hb^S em diferentes regiões surgiram independentemente, não por meio da difusão por migração. Na verdade, com o advento de ferramentas para a genotipagem do DNA, está claro que a mutação Hb^S surgiu independentemente em cinco haplótipos diferentes e sua frequência aumentou em certas regiões. Com base na diversidade genética limitada das populações com malária, acredita-se que a mutação Hb^S tenha surgido nos últimos séculos, após as populações começarem a viver ao redor de cursos de água com o advento da agricultura.

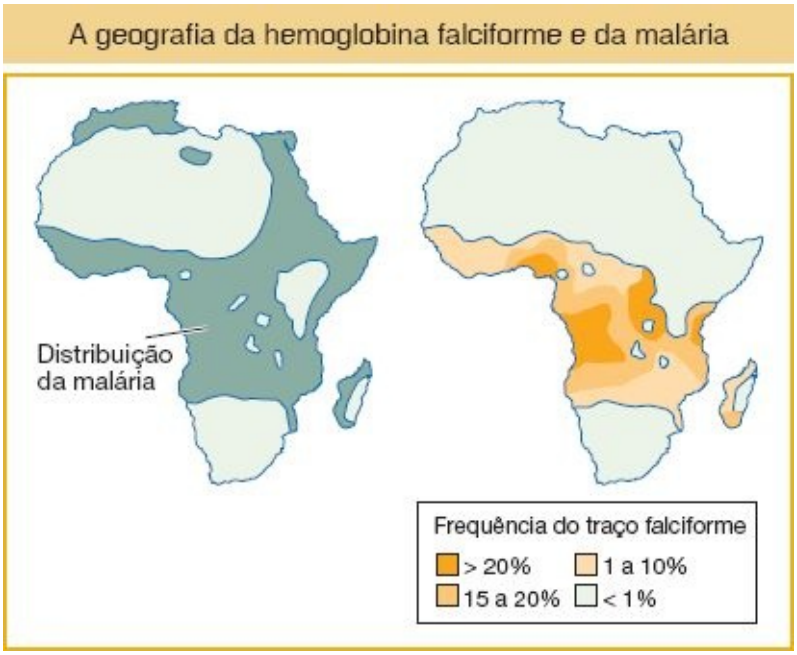


FIGURA 20.6 Estes mapas demonstram a correspondência próxima entre a distribuição da malária (*esquerda*) e a frequência do traço falciforme (*direita*) na África. (Dados de A. C. Allison, *Genetics* 66, 2004, 1591; redesenhada por Leanne Olds.)

> CONCEITO-CHAVE O papel da mutação da hemoglobina *S* falciforme de conferir resistência à malária foi o primeiro exemplo de seleção natural a ser elucidado, no qual o agente de seleção foi demonstrado, o valor adaptativo relativo dos diferentes genótipos pôde ser medido e a base genética e molecular da variação funcional foi revelada.

O papel da Hb^S de conferir resistência à malária ilustrou três facetas importantes do processo evolutivo:

1. *A evolução pode se repetir, e se repete*. As múltiplas origens e expansões independentes da mutação Hb^S demonstram que quando são proporcionados um tamanho de população e um tempo suficiente, as mesmas mutações podem surgir e se difundir repetidamente. Atualmente, são conhecidos muitos outros exemplos de repetição independente e precisa da evolução de mutações adaptativas, e encontraremos diversos outros neste capítulo.
2. *A adaptabilidade é um estado condicional muito relativo*. Se uma mutação é vantajosa, desvantajosa ou nenhuma das duas depende muito das condições ambientais. Na ausência de malária, Hb^S é muito rara e desfavorecida. Onde a malária está presente, ela pode alcançar altas frequências, apesar das desvantagens impostas aos homozigotos *SS*. Em afro-americanos, a frequência de Hb^S está em declínio em virtude da seleção contra o alelo na ausência de malária na América do Norte.
3. *A seleção natural atua sobre qualquer variação que esteja disponível, não necessariamente pelos melhores meios imagináveis*. A mutação Hb^S, embora seja uma proteção contra a malária, também causa uma condição de risco à vida. Nas áreas em que a malária é prevalente, onde vivem mais de 40% da população mundial, o imperativo de combater a malária contrabalanceia o efeito deletério da mutação falciforme.

20.3 Evolução molecular | Teoria neutra

Darwin e Wallace conceberam a evolução amplamente como “alterações nos organismos ocasionadas por seleção natural”. Na verdade, é assim que a maior parte das pessoas pensa a respeito do significado de “evolução”. Entretanto, um século após a teoria de Darwin, na medida em que biólogos moleculares começaram a confrontar a evolução no nível das proteínas e das moléculas de DNA, eles encontraram e identificaram outra dimensão do processo evolutivo, a evolução molecular neutra, que não envolve a seleção natural. Uma compreensão sobre a evolução molecular neutra é crucial para entender como os genes se alteram ao longo do tempo.

Desenvolvimento da teoria neutra

Na década de 1950 e no início da década de 1960, foram desenvolvidos métodos que possibilitaram que os biólogos determinassem as sequências de aminoácidos das proteínas. Essa nova capacidade deu origem à perspectiva de que a base fundamental da alteração evolutiva finalmente estivesse ao alcance das mãos. Entretanto, na medida em que as sequências de proteínas de uma diversidade de espécies foram decifradas, surgiu um paradoxo. As sequências das globinas e do citocromo c, por exemplo, diferem tipicamente entre duas espécies quaisquer no número de aminoácidos, e esse número aumenta com o tempo decorrido desde a sua divergência a partir de um ancestral comum (Figura 20.7). Ainda assim, a função dessas proteínas é a mesma em diferentes espécies – transportar e disponibilizar oxigênio para os tecidos no caso da hemoglobina, e alternar elétrons durante a respiração celular no caso do citocromo c.

O quebra-cabeça era então se as substituições de aminoácidos entre as espécies refletiam alterações na função das proteínas e adaptações às condições seletivas. Os bioquímicos Linus Pauling e Emile Zuckerkandl não achavam isso. Eles observaram que muitas substituições eram de um aminoácido por outro com propriedades semelhantes. Eles concluíram que a maior parte das substituições de aminoácidos era “neutra” ou “quase neutra” e que elas não alteravam em nada a função de uma proteína.

Essa linha de raciocínio foi rejeitada à primeira vista por muitos biólogos evolutivos, que na ocasião consideravam todas as alterações evolutivas resultado

da seleção natural e da adaptação. O paleontólogo George Gaylord Simpson argumentou que “existe um forte consenso de que genes ou alelos completamente neutros devem ser raros, se chegarem a existir. Para um biólogo evolutivo, portanto, parece ser altamente improvável que as proteínas (…) possam ser alteradas de maneira regular, porém não adaptativa”.[6]

Zuckerkandl e Pauling afirmaram que a semelhança ou as diferenças entre os organismos não precisam ser refletidas no nível da proteína – que a alteração molecular e a alteração visível não eram necessariamente ligadas ou proporcionais.

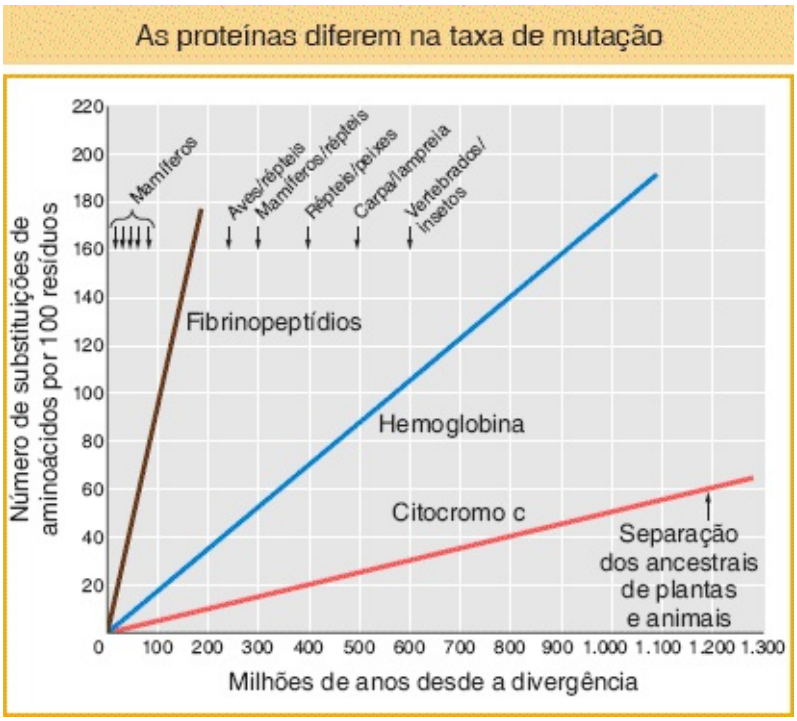


FIGURA 20.7 Número de substituições de aminoácidos na evolução dos vertebrados como uma função do tempo desde a divergência. As três proteínas – fibrinopeptídios, hemoglobina e citocromo c – diferem na taxa de substituições em virtude de as diferentes proporções de suas substituições de aminoácidos serem seletivamente neutras.

O debate foi resolvido por um ataque de dados empíricos e pela elucidação do código genético. Tendo em vista que múltiplos códons codificam o mesmo aminoácido, uma mutação que altera, digamos, CAG para CAC, não altera o aminoácido codificado. Portanto, pode existir uma variação no nível do DNA que

não apresenta efeitos sobre as sequências proteicas e, portanto, existem alelos neutros. No entanto, mais importante para a genética de populações foi o desenvolvimento da "teoria neutra da evolução molecular" por Motoo Kimura, Jack L. King e Thomas Jukes. Esses autores propuseram que a maior parte das mutações que são fixadas é neutra ou quase neutra, mas não todas, e que quaisquer diferenças entre as espécies em tais sítios no DNA evoluem por deriva genética aleatória.

A "teoria neutra" marcou uma profunda alteração conceitual, que se distancia da consideração sobre a evolução sempre orientada por seleção natural. Além disso, ela proporcionou uma base sobre como o DNA deve mudar ao longo do tempo se nenhum outro agente, tal como a seleção natural, intervier.

> CONCEITO-CHAVE A teoria neutra da evolução molecular propôs que a maior parte das mutações no DNA ou das substituições de aminoácidos entre as espécies é funcionalmente neutra ou quase neutra e fixada por deriva genética aleatória. A suposição de neutralidade oferece uma expectativa basal sobre como o DNA deve mudar ao longo do tempo quando a seleção natural está ausente.

Taxa de substituições neutras

Conforme vimos no Capítulo 18 (ver Quadro 18.5), podemos calcular a taxa esperada de alterações neutras nas sequências de DNA ao longo do tempo. Se μ é a taxa de novas mutações em um *locus* por cópia de gene por geração, o número absoluto de novas mutações que aparecerão em uma população de N indivíduos diploides é, então, $2N\mu$. As novas mutações estão sujeitas à deriva genética aleatória: a maior parte será perdida da população, enquanto algumas se tornarão fixas e substituirão o alelo anterior. Se uma mutação que se originou recentemente for neutra, existe então uma probabilidade de $1/(2N)$ de que ela substitua o alelo anterior em virtude da deriva genética aleatória. Cada uma das $2N\mu$ novas mutações que aparecerão em uma população apresenta uma probabilidade de

1/(*2N*) de finalmente fazer parte dessa população. Portanto, a taxa de substituição absoluta *k* é a taxa de mutação multiplicada pela probabilidade de que qualquer mutação acaba sofrendo deriva:

$$K = \text{Taxa de substituição neutra} = 2N\mu \times 1/(2N) = \mu$$

Ou seja, esperamos que, a cada geração, ocorram μ substituições na população, puramente a partir de deriva genética das mutações neutras.

> CONCEITO-CHAVE A taxa de substituições no DNA na evolução que resulta de deriva genética aleatória das mutações neutras é igual à taxa de mutação para tais alelos, μ.

Assinatura da seleção purificadora no DNA

Quando as medidas da alteração molecular se desviam do que é esperado em relação às alterações neutras, isso é um sinal importante – um sinal de que a seleção interveio. Tal sinal pode revelar que a seleção favoreceu alguma alteração específica ou que rejeitou outras. Vimos, no caso da mutação Hb^S, como a seleção natural favorece a mutação na presença do parasita da malária, mas a rejeita onde a malária está ausente. A influência mais pervasiva da seleção natural sobre o DNA, na verdade é conservar a função e a sequência do gene.

Todas as classes de sequências de DNA, incluindo éxons, íntrons, sequências reguladoras e sequências entre genes, demonstram diversidade de nucleotídios entre os indivíduos nas populações e entre as espécies. A taxa constante de substituições neutras prevê que, se o número de diferenças nucleotídicas entre duas espécies for inserido em um gráfico em face do tempo desde a sua divergência de um ancestral comum, o resultado deverá ser uma linha reta, com uma inclinação igual a μ. Ou seja, a evolução deve avançar de acordo com um **relógio molecular** que está operando a uma taxa μ. A Figura 20.8 demonstra um gráfico em relação ao gene da β-globina. Os resultados são consideravelmente consistentes com a alegação de que as substituições de nucleotídios têm sido

neutras nos últimos 500 milhões de anos. Dois tipos de substituições de nucleotídios neutras estão inseridos no gráfico: **substituições sinônimas**, que são de um códon alternativo para outro, sem alterar o aminoácido, e **substituições não sinônimas**, que resultam em alteração no aminoácido. A Figura 20.8 demonstra uma inclinação muito mais baixa em relação às substituições não sinônimas do que em relação às substituições sinônimas, o que significa que a taxa de substituição das substituições não sinônimas neutras é muito mais baixa do que aquela das substituições neutras sinônimas.

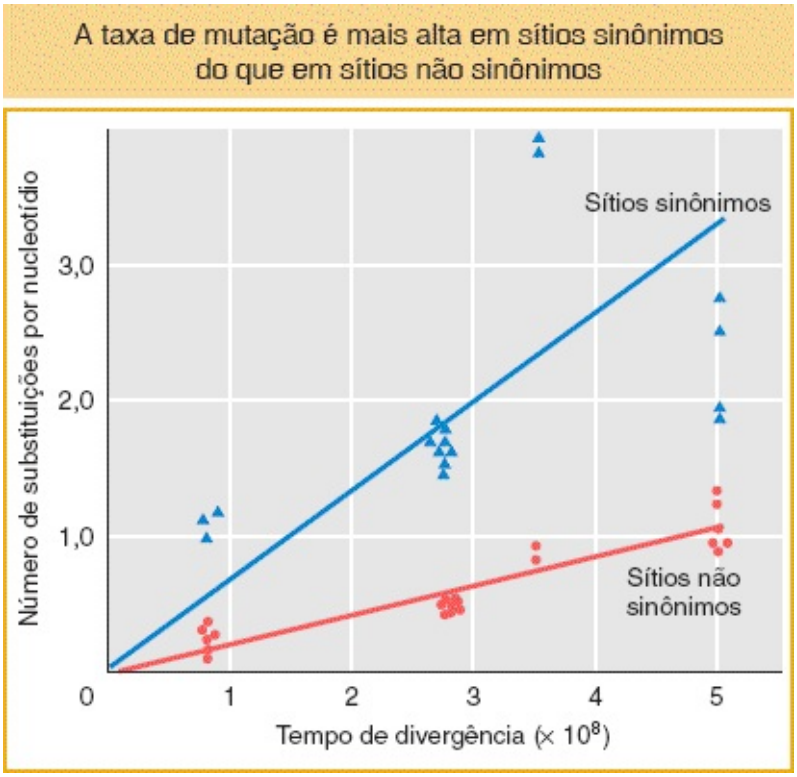


FIGURA 20.8 A divergência nucleotídica em sítios sinônimos é maior do que em sítios não sinônimos do gene da β-globina.

Esse desfecho é precisamente o que esperamos sob seleção natural. As mutações que causam a substituição de um aminoácido devem apresentar um efeito deletério com mais frequência do que as substituições sinônimas, que não alteram a proteína. As variantes deletérias serão removidas das populações pela *seleção purificadora* (ver Capítulo 18). Uma razão inferior à esperada entre

alterações não sinônimas e sinônimas é uma assinatura da seleção purificadora. É importante notar que essas observações não demonstram que as substituições sinônimas não sofrem restrições seletivas sobre elas. Em vez disso, elas demonstram que essas restrições, em média, não são tão fortes quanto aquelas das mutações que alteram aminoácidos. Assim, uma alteração sinônima, ainda que não apresente efeito sobre a sequência de aminoácidos, altera o mRNA naquela sequência e, portanto, pode afetar a estabilidade do mRNA ou a eficiência em que o mRNA é traduzido.

A seleção purificadora é a faceta mais disseminada, mas frequentemente negligenciada da seleção natural. A "rejeição de variações danosas", tal como Darwin denominou, é universal. A seleção purificadora explica o motivo de encontrarmos tantas sequências de proteínas que são inalteradas ou quase inalteradas ao longo de vastos períodos de tempo evolutivo. Por exemplo, há várias dezenas de genes que existem em todos os domínios da vida – Archaea, bactérias, fungos, plantas e animais – que codificam proteínas cujas sequências têm sido amplamente conservadas ao longo de 3 bilhões de anos de evolução. Para preservar as referidas sequências, variantes que surgiram aleatoriamente em bilhões de indivíduos em dezenas de milhões de espécies foram rejeitadas por seleção repetidamente.

> CONCEITO-CHAVE A seleção purificadora é um aspecto penetrante da seleção natural, que reduz a variação genética e preserva as sequências de DNA e proteínas por muito tempo.

Outra previsão da teoria da evolução neutra é que diferentes proteínas apresentarão diferentes taxas temporais, tendo em vista que as funções metabólicas de algumas proteínas serão muito mais sensíveis às alterações em suas sequências de aminoácidos. As proteínas nas quais cada aminoácido faz uma diferença apresentarão uma taxa mais baixa de mutação neutra, tendo em vista que uma menor proporção de suas mutações será neutra em comparação às proteínas que são mais tolerantes à substituição. A Figura 20.7 demonstra uma comparação

dos relógios para os fibrinopeptídios, a hemoglobina e o citocromo c. É razoável que os fibrinopeptídios apresentem uma proporção muito mais alta de mutações neutras, tendo em vista que esses peptídios são meramente uma trava de segurança não metabólica, retirada do fibrinogênio para ativar a reação de coagulação sanguínea. Não está claro o motivo pelo qual as hemoglobinas são menos sensíveis às alterações de aminoácidos do que o citocromo c.

> CONCEITO-CHAVE A taxa de evolução neutra em relação à sequência de aminoácidos de uma proteína depende da sensibilidade da função da proteína às alterações de aminoácidos.

A conservação das sequências gênicas por meio da seleção purificadora e a evolução neutra de sequências gênicas são duas dimensões cruciais do processo evolutivo, mas nenhuma delas explica a origem das adaptações. Nas duas próximas seções, ilustraremos diversos exemplos dos modos por meio dos quais as alterações genéticas estão ligadas às alterações nos traços e à diversidade dos organismos.

20.4 Seleção cumulativa e vias de múltiplas etapas até a alteração funcional

Tendo em vista que uma grande parte da evolução é neutra, não existe uma relação simples entre a quantidade de alterações no DNA de um gene e a quantidade de alterações, se existente, na função da proteína codificada. Em um extremo, quase toda a sequência de aminoácidos de uma proteína pode ser substituída ao mesmo tempo que se mantém a função original, se aqueles aminoácidos que são substituídos mantiverem a estrutura tridimensional da enzima.

Contrariamente, a função de uma enzima pode ser alterada pela substituição de um único aminoácido. A mosca dos ovinos, *Lucilia cuprina*, desenvolveu

resistência a inseticidas organofosforados amplamente utilizados para o seu controle. Richard Newcombe, Peter Campbell e seus colegas demonstraram que essa resistência é a consequência de uma única substituição do ácido aspártico por um resíduo de glicina no sítio ativo de uma enzima que é ordinariamente uma carboxilesterase (desmembra um éster carboxílico, R-COO-R, em um álcool e um carboxilato). A mutação causa a perda completa da atividade de carboxilesterase e a sua substituição pela atividade de esterase (desmembra qualquer éster, R-O-R, em um ácido e um álcool). A modelagem tridimensional da molécula indica que a proteína substituída ganha a capacidade de se ligar a uma molécula de água próxima do sítio de ligação do organoforado. Em seguida, a molécula de água reage com o organoforado, dividindo-o em dois.

> CONCEITO-CHAVE Não existe uma relação proporcional entre quanta alteração ocorre no DNA na evolução e quanta alteração funcional resulta.

A seleção claramente desempenha um papel na evolução da carboxilesterase do inseto e na resistência ao inseticida. Entretanto, em muitos casos, as substituições de aminoácidos que alteram a função da proteína são mais numerosas e se acumulam durante repetidas rodadas de mutação e seleção, ao que se denomina **seleção cumulativa**. O poder da seleção cumulativa de direcionar alterações maiores na função de uma molécula é uma das facetas menos apreciadas da evolução por seleção natural. Um motivo é que o papel da seleção em cada uma das múltiplas substituições é mais difícil de certificar.

> CONCEITO-CHAVE A seleção cumulativa pode impulsionar a fixação de muitas alterações em moléculas em evolução.

Com a finalidade de compreender o papel da seleção em casos de substituições múltiplas, são adotadas duas abordagens principais: análise experimental empírica e métodos estatísticos. Ilustraremos, de início, a primeira delas.

Vias de múltiplas etapas na evolução

Quando as mutações surgem em múltiplos sítios na evolução de um estado fenotípico para outro, existem múltiplas ordens possíveis nas quais essas mutações podem aparecer, cada uma representando uma via diferente no espaço genético que a evolução pode assumir. Tais vias de múltiplas etapas da alteração evolutiva são denominadas **vias adaptativas**.

Suponha que a diferença entre o fenótipo original e a forma que evoluiu seja uma consequência de mutações em cinco sítios, A, B, C, D e E. Existem muitas ordens diferentes nas quais essas mutações podem ter ocorrido ao longo da evolução. Primeiramente, o sítio A pode ter sido fixado na população, em seguida D, em seguida C, em seguida E e, finalmente, B. Por outro lado, a ordem de fixação pode ter sido E, D, A, B, C. Em relação a cinco sítios, existem $5 \times 4 \times 3 \times 2 \times 1 = 120$ ordens possíveis. Duas questões importantes na compreensão da evolução são: "Quantas dessas vias evolutivas alternativas são possíveis?" e "Quais são as probabilidades de diferentes vias possíveis em relação umas às outras?"

Daniel Weinreich e seus colegas caracterizaram em detalhes tal conjunto de vias adaptativas no espaço genético em seu estudo sobre a evolução da resistência a antibiótico na bactéria *E. coli*. A resistência ao antibiótico cefotaxima é adquirida por meio do acúmulo de cinco mutações em diferentes sítios no gene da β-lactamase bacteriana. Quatro das mutações levam a alterações de aminoácidos e a quinta é uma mutação não codificadora. Quando todas as cinco mutações estão presentes, a concentração mínima de antibiótico necessária para inibir o crescimento bacteriano aumenta em um fator de 100.000. Os experimentadores primeiramente mediram a resistência conferida por uma mutação em um determinado sítio na presença de todas as $2^4 = 16$ combinações possíveis de mutantes e não mutantes nos outros quatro sítios. Na maior parte das combinações, mas não em todas, um mutante em um sítio era mais resistente, independentemente do estado dos outros quatro sítios. Por exemplo, um mutante no sítio G238S demonstrou resistência significativa, independentemente do estado mutante ou não mutante dos outros quatro sítios (Tabela 20.3). Por outro lado, a

mutação no sítio não codificador g4205a conferiu resistência significativa em oito combinações, alteração insignificante na resistência em seis casos e uma *diminuição* na resistência em duas combinações. Essa dependência da vantagem ou da desvantagem adaptativa de uma nova mutação sobre as mutações que haviam sido fixadas anteriormente é o que os experimentadores denominam **epistasia de sinal**.

Weinrich e colegas mediram a resistência a cada etapa na sequência temporal da adição de mutações, um sítio após o outro. Se uma mutação em uma das 120 ordens possíveis não conferisse uma resistência mais alta, presumivelmente aquela via evolutiva seria então encerrada, tendo em vista que não haveria seleção em favor da mutação ou mesmo contra ela. Eles observaram que, das 120 vias possíveis pelo histórico da mutação, apenas 18 conferiam aumento de resistência a cada etapa da mutação. Portanto, 102/120 = 85% das possíveis vias de mutação até a resistência máxima não eram acessíveis para a evolução por seleção natural. Finalmente, presumimos que, em uma população que desenvolve resistência, a probabilidade de que uma via acessível em particular de fato seja seguida é proporcional à magnitude do aumento da resistência a cada etapa. Sob essa premissa, apenas 10 das 18 vias acessíveis explicarão 90% dos casos de evolução de resistência bacteriana ao antibiótico (Figura 20.9).

> CONCEITO-CHAVE A ordem na qual as mutações ocorrem é de importância crítica para determinar a via de evolução e se a evolução por seleção natural irá ou não de fato alcançar o estado mais vantajoso. Tendo em vista que a ordem de ocorrência das mutações é aleatória, muitos fenótipos vantajosos podem nunca ser alcançados, muito embora ocorram mutações individuais.

Tabela 20.3 A dependência dos efeitos da adaptabilidade de mutações sobre mutações anteriores em *E. coli*.

Número de alelos nos quais o efeito mutacional médio é

Mutação*	Positivo	Negativo	Insignificante	Aumento proporcional médio
g4205a	8	2	6	1,4
A42G	12	0	4	5,9
E104K	15	1	0	9,7
M182T	8	3	5	2,8
G238S	16	0	0	$1{,}0 \times 10^{3}$

*As mutações que levam à resistência ao antibiótico são designadas pela posição de seus nucleotídios ou aminoácidos. Das 16 possíveis combinações alélicas dos quatro outros sítios, os efeitos positivos, negativos ou neutros da mutação estão indicados juntamente com o aumento proporcional médio na adaptabilidade em relação às mutações no sítio indicado.

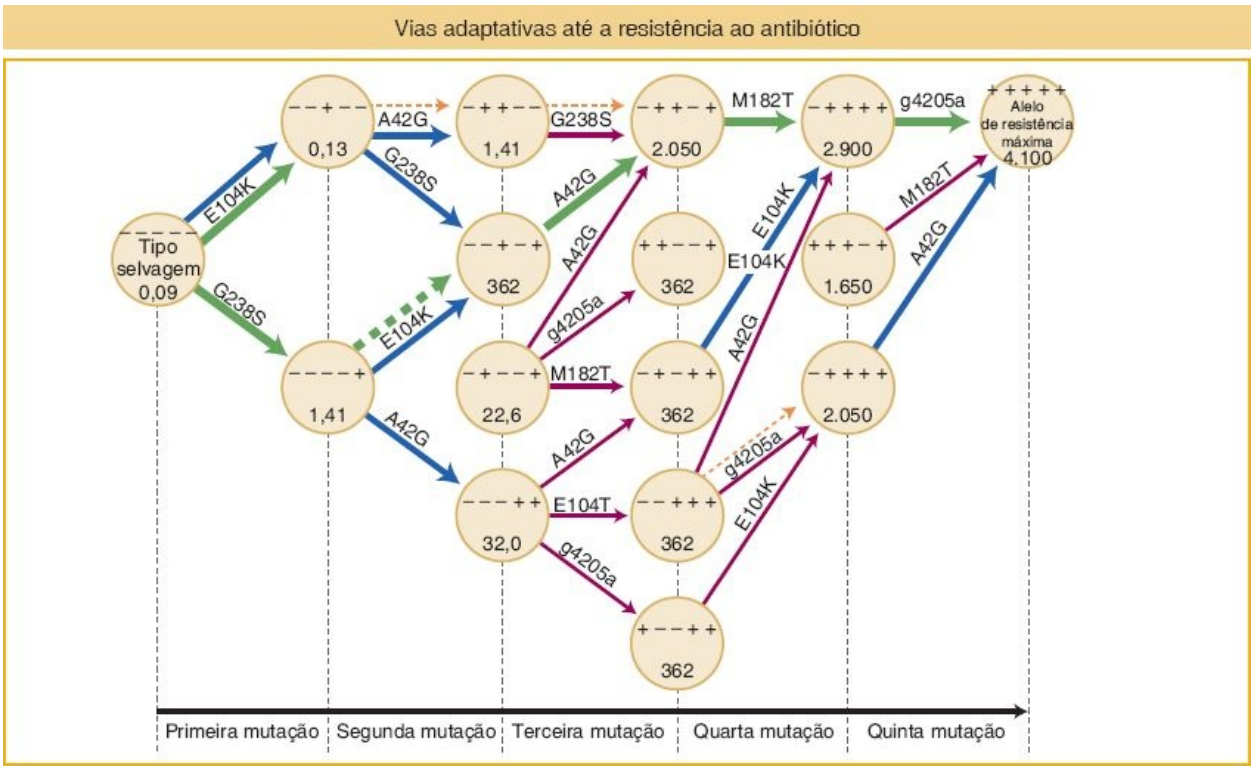


FIGURA 20.9 As etapas mutacionais em relação às 10 trajetórias mais prováveis desde a suscetibilidade do tipo selvagem ao antibiótico cefotaxima até a resistência máxima. Cada círculo representa um alelo cuja identidade é indicada por um trecho de cinco símbolos de + ou –, correspondentes (*esquerda para a direita*) à presença ou à ausência das mutações g4205a, A42G, E104K, M182T e G238S, respectivamente. Os números indicam o grau de resistência à cefotaxima em microgramas por mililitro. A probabilidade relativa de cada mutação benéfica é representada pela cor e pela largura das setas: *verde* – larga, mais alta; *azul* – média, moderada; *roxa* – estreita, baixa; e *laranja* – muito estreita, a mais baixa. (Dados de D. Weinreich *et al.*, *Science* 312, 2006, 111-114.)

Então, um fator-chave na determinação da via evolutiva que uma população pode seguir é a aleatoriedade do processo de mutação. Após a variação genética inicial ser exaurida pela fixação seletiva e aleatória de alelos, a nova variação que tem origem a partir da mutação pode ser a fonte da alteração evolutiva adicional. A direção em particular dessa evolução adicional depende das mutações em particular que ocorrem e da ordem temporal na qual elas surgem.

Uma ilustração muito clara dessa contingência histórica de vias adaptativas é um experimento de seleção realizado por Holly Wichman e seus colegas. Eles forçaram o bacteriófago ΦX174 a se reproduzir em altas temperaturas e no hospedeiro *Salmonella typhimurium*, em vez de no seu hospedeiro normal,

Escherichia coli. Duas linhagens independentes de vírus foram estabelecidas, rotuladas TX e ID, e mantidas em separado, embora ambas tenham sido expostas às mesmas condições. Ambas desenvolveram a capacidade de se reproduzir em altas temperaturas no novo hospedeiro. Em uma das duas linhagens, a capacidade de se reproduzir na *E. coli* ainda existia, mas, na outra linhagem, a capacidade foi perdida. O bacteriófago apresenta apenas 11 genes e, assim, os experimentadores puderam registrar as sucessivas alterações no DNA em relação a todos esses genes e nas proteínas codificadas por eles durante o processo de seleção. Houve 15 alterações do DNA na linhagem TX, localizadas em seis genes diferentes; na linhagem ID, houve 14 alterações localizadas em quatro genes diferentes. Em sete casos, as alterações nas duas linhagens foram idênticas, incluindo uma grande deleção, mas até mesmo essas alterações idênticas apareceram em cada linhagem em uma ordem diferente (Tabela 20.4). Assim, por exemplo, a alteração no sítio 1533 do DNA, que causou uma substituição de isoleucina por treonina, foi a terceira alteração na linhagem ID, mas a 14ª alteração na linhagem TX.

Portanto, o curso da evolução seguido pelos vírus inicialmente idênticos dependeu das mutações disponíveis em um dado momento no processo de seleção cumulativa. Contraste essa situação com a origem repetida do alelo falciforme Hb^S: nesse caso, a mesma mutação surgiu e se difundiu cinco vezes. Claramente, em alguns casos existem muitas "soluções" moleculares para as condições seletivas e, em outros, apenas uma ou poucas.

> CONCEITO-CHAVE Sob condições idênticas de seleção natural, duas populações podem chegar a composições genéticas idênticas ou diferentes como resultado direto da seleção natural.

Tabela 20.4 Substituições moleculares em dois bacteriófagos Φ-X174, TX e ID, durante adaptação.

Ordem*	Sítio TX	Troca de aminoácido	Sítio ID	Troca de aminoácido

1	782	E72, T → 1	2167	F388, H → Q
2	1727	F242, L → F	1613	F204, T → S
3	2085	F361, A → V	**1533**[6]	F177, T → I
4	319	C63, V → F	1460	F153, Q → E
5	2973	H15, G → S	1300	F99, silenciosa
6	323	C64, D → G	**1305**[3]	F101, G → D
7	**4110**[3]	A44, H → Y	1308	F102, Y → C
8	1025	F8, E → K	4110[1]	A44, H → Y
9	**3166**[7]	H79, A → V	4637	A219, silenciosa

10	5185	A402, T → M	**965-91**[4]	deleção
11	**1305**[2]	F101, G → D	**5365**[5]	A462, M → T
12	**965-91**[4]	deleção	**4168**[7]	A63, Q → R
13	**5365**[5]	A462, M → T	**3166**[2]	H79, A → V
14	**1533**[1]	F177, T → I	1809	F269, K → R
15	**4168**[6]	A63, Q → R		

*As alterações estão listadas na ordem em que apareceram em cada uma das duas linhagens de bacteriófagos durante a seleção. A posição do nucleotídio está listada, seguida pela proteína afetada, A-H, com o número do resíduo de aminoácido e a natureza da substituição do aminoácido. As alterações em paralelo estão demonstradas em negrito, e um sobrescrito indica a ordem daquelas alterações na outra linhagem viral de seleção.
Fonte: H. A. Wichman *et al*. *Science* 285, 1999, 422-424.

A dissecção experimental das vias evolutivas consome muito tempo e é dispendiosa. Além disso, com frequência não é prático para os experimentadores modificar todo genótipo possível em uma via adaptativa nas populações, ou tentar medir a adaptabilidade relativa de muitos organismos na natureza. Os exemplos de resistência a antibióticos e hospedeiros virais são casos em que tanto a modificação genética quanto medições de adaptabilidade são prontamente

executadas em bactérias e seus vírus em laboratório. Em outras situações, foram planejados métodos estatísticos para revelar uma assinatura que indicasse que a seleção atuou nas sequências de DNA e proteínas.

Assinatura de seleção positiva nas sequências de DNA

A demonstração do relógio molecular argumenta que a maior parte das substituições de nucleotídios que ocorreram na evolução foi neutra, mas não nos informa o quanto da evolução molecular consistiu em uma alteração adaptativa direcionada por seleção positiva. Um modo de detectar a evolução adaptativa de uma proteína é por meio da comparação dos polimorfismos sinônimos e não sinônimos de nucleotídios dentro das espécies com as alterações sinônimas e não sinônimas nos nucleotídios entre espécies. Se todas as mutações forem neutras, a razão de polimorfismos não sinônimos de nucleotídios em relação aos sinônimos dentro de uma espécie deverá ser a mesma que a razão de substituições de nucleotídios não sinônimos em relação aos sinônimos entre espécies. Por outro lado, se as alterações de aminoácidos entre as espécies houverem sido direcionadas por seleção positiva, deve haver um excesso de alterações não sinônimas entre as espécies.

Um teste para detectar a seleção positiva nas sequências de DNA foi desenvolvido por John McDonald e Martin Kreitman. Esse teste envolve diversas etapas lógicas, porém simples:

1. A sequência de DNA de um gene é obtida a partir de um número de indivíduos ou linhagens em separado de cada uma das duas espécies. Dez ou mais sequências de cada espécie seriam desejáveis. Em seguida, as diferenças de nucleotídios fixados entre as espécies são classificadas em diferenças não sinônimas (*a*) e sinônimas (*b*).
2. As diferenças de nucleotídios entre os indivíduos dentro de cada espécie (polimorfismos) são tabuladas em seguida, e classificadas como aquelas que resultam em mudanças de aminoácidos (polimorfismos não sinônimos; *c* na tabela a seguir), ou aquelas que não alteram o aminoácido (polimorfismos sinônimos; *d* na tabela a seguir).

3. Se a divergência entre as espécies for puramente o resultado de deriva genética aleatória, esperamos então que *a/b* seja igual a *c/d*. Se, por outro lado, houver ocorrido uma divergência seletiva, deverá haver um excesso de diferenças não sinônimas fixadas e, assim, *a/b* deve ser maior que *c/d*.

	Diferenças fixadas nas espécies	Polimorfismos
Não sinônimos	*a*	*c*
Sinônimos	*b*	*d*
Razão	*a/b*	*c/d*

A Tabela 20.5 demonstra uma aplicação desse princípio para o gene da desidrogenase alcoólica em três espécies de *Drosophila* relacionadas de modo próximo. Claramente, existe um excesso de substituições de aminoácidos entre as espécies acima do que é esperado. Portanto, concluímos que algumas das substituições de aminoácidos na enzima foram alterações adaptativas direcionadas por seleção natural.

Tabela 20.5 Polimorfismos sinônimos e não sinônimos e diferenças em relação à desidrogenase alcoólica em três espécies de *Drosophila*.

Organismo	**Diferenças nas espécies**	**Polimorfismos**
Não sinônimo	7	2
Sinônimo	17	42

Razão	0,29:0,71	0,05:0,95

Fonte: J. McDonald e M. Kreitman, “Adaptive Protein Evolution at the *Adh* locus in *Drosophila*”, *Nature* 351, 1991, 652-654.

20.5 Evolução morfológica

Uma das categorias mais óbvias e interessantes da evolução dos traços é aquela da morfologia dos organismos. Entre animais, por exemplo, existe uma grande diversidade no número, tipo, tamanho, formato e cor das partes corporais. Tendo em vista que a forma adulta é o produto do desenvolvimento embrionário, as alterações na forma devem ser o resultado de alterações que ocorrem durante o desenvolvimento. Avanços recentes na compreensão do controle genético do desenvolvimento (ver Capítulo 13) possibilitaram que os pesquisadores investigassem as bases genéticas e moleculares da evolução da forma dos animais. Veremos que algumas das alterações dramáticas na forma dos animais apresentam uma base genética e molecular relativamente simples, enquanto a evolução dos traços regulados por muitos genes *toolkit* envolve mecanismos moleculares que são distintos daqueles que examinamos até agora. Examinaremos casos nos quais as substituições codificadoras, a inativação de genes e a evolução da sequência reguladora, respectivamente, são a base da divergência morfológica.

Alterações adaptativas em uma proteína reguladora de pigmento

Alguns dos exemplos mais surpreendentes e mais bem-compreendidos da divergência morfológica são encontrados nos padrões da coloração corporal em animais. A pelagem dos mamíferos, a plumagem das aves, as escamas dos peixes e os esquemas de cores das asas dos insetos são maravilhosamente diversos. Investigadores realizaram muitos progressos na compreensão sobre o controle genético da formação das cores e de seu papel na evolução das diferenças de colorações dentro das espécies e entre elas.

Na região de Pinacate no sudoeste do Arizona, afloramentos rochosos escuros são circundados por granito arenoso de cor mais clara (Figura 20.10). O pequeno

roedor *Chaetodipus intermedius* habita Pinacate, bem como outras áreas rochosas dessa região. Os camundongos encontrados nas rochas magmáticas são tipicamente de coloração escura, enquanto aqueles encontrados nas áreas adjacentes de granito de cor arenosa ou no solo desértico em geral são normalmente de coloração clara (Figura 20.11). Estudos de campo sugerem que a correspondência entre a cor da coloração da pelagem e do ambiente protege os camundongos contra a sua visualização por parte de predadores.

O roedor *Chaetodipus intermedius* fornece um exemplo do *melanismo* – a ocorrência de uma forma escura em uma população ou espécie. O melanismo é um dos tipos mais comuns de variação fenotípica em animais. A coloração escura da pelagem ocorre em virtude da pesada deposição do pigmento *melanina*, o pigmento mais difundido no reino animal. Em mamíferos, dois tipos de melanina são produzidos nos melanócitos (as células de pigmento da epiderme e dos folículos pilosos): eumelanina, que forma os pigmentos pretos ou marrons, e feomelanina, que forma os pigmentos amarelos ou vermelhos. As quantidades relativas de eumelanina e feomelanina são controladas pelos produtos de diversos genes. Duas proteínas-chave são o receptor de melanocortina 1 (MC1R) e a proteína aguti. Durante o ciclo de crescimento dos pelos, o hormônio estimulante dos melanócitos α (α-MSH) se liga à proteína MC1R, que aciona a indução das enzimas produtoras dos pigmentos. A proteína aguti bloqueia a ativação de MC1R e inibe a produção de eumelanina.

Cores contrastantes do deserto de Pinacate

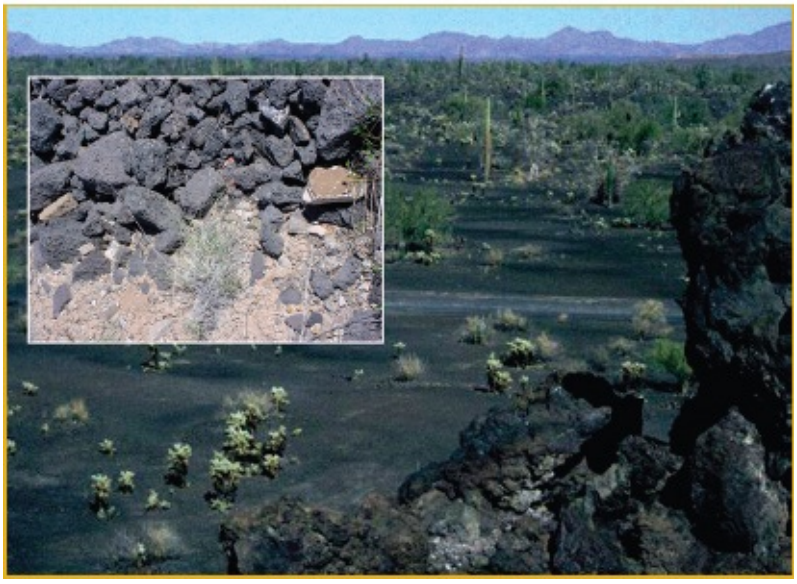


FIGURA 20.10 Fluxos de lava no deserto de Pinacate produziram rochas pretas adjacentes aos substratos de cor arenosa. (Michael Nachman, University of Arizona.)

Melanismo no camundongo *rock pocket*

FIGURA 20.11 *Chaetodipus intermedius* de coloração clara e escura da região de Pinacate, no Arizona, estão demonstrados em solo com cor arenosa e na rocha magmática preta. (Michael Nachman, de M. W. Nachman *et al.*, "The Genetic Basis of Adaptive Melanism in Pocket Mice", *Proc. Natl. Acad. Sci. USA* 100, 2003, 5268-5273.)

Michael Nachman e seus colegas examinaram as sequências de DNA dos genes *mcr1* dos roedores *Chaetodipus intermedius* de coloração clara e escura. Eles observaram a presença de quatro mutações no gene *mc1r* em roedores escuros, que fazem com que a proteína MC1R seja diferente em quatro resíduos de aminoácidos da proteína correspondente em roedores claros. Achados de estudos bioquímicos sugerem que tais mutações fazem com que a proteína MC1R seja constitutivamente ativa (ativa em todas as ocasiões), evitando (*bypassing*) a regulação da atividade receptora pela proteína aguti. Na verdade, mutações em *mc1r* estão associadas ao melanismo em todos os tipos de vertebrados, selvagens e domesticados. Muitas dessas mutações alteram resíduos na mesma parte da proteína MC1R e as mesmas mutações ocorreram independentemente em algumas espécies (Figura 20.12).

De muitas maneiras, podemos pensar a respeito desses roedores escuros como análogos aos tentilhões das Galápagos, e das rochas magmáticas como novos *habitats* "insulares" produzidos pela mesma atividade vulcânica que produziu as ilhas Galápagos. O tipo de coloração arenosa do roedor aparenta ser o tipo ancestral, semelhante ao tentilhão continental ancestral que colonizou Galápagos. A vantagem de ser menos visível para os predadores resultou em seleção natural em relação à coloração da pelagem e a invasão das ilhas de rochas magmáticas pelos roedores levou à difusão de um alelo que foi favorecido no solo rochoso negro e selecionado contra o solo de coloração arenosa. Novas mutações no gene *mc1r* foram essenciais para essa adaptação ao cenário em alteração.

A evolução do melanismo em roedores *Chaetodipus intermedius* ilustra como a adaptabilidade depende das condições nas quais um organismo vive. A nova mutação preta foi favorecida no ambiente com rochas magmáticas, mas desfavorecida na população ancestral que vive em terreno de cor arenosa.

> CONCEITO-CHAVE A adaptabilidade relativa de uma nova variante depende das condições seletivas imediatas. Uma mutação que pode ser benéfica em uma população pode ser deletéria em outra.

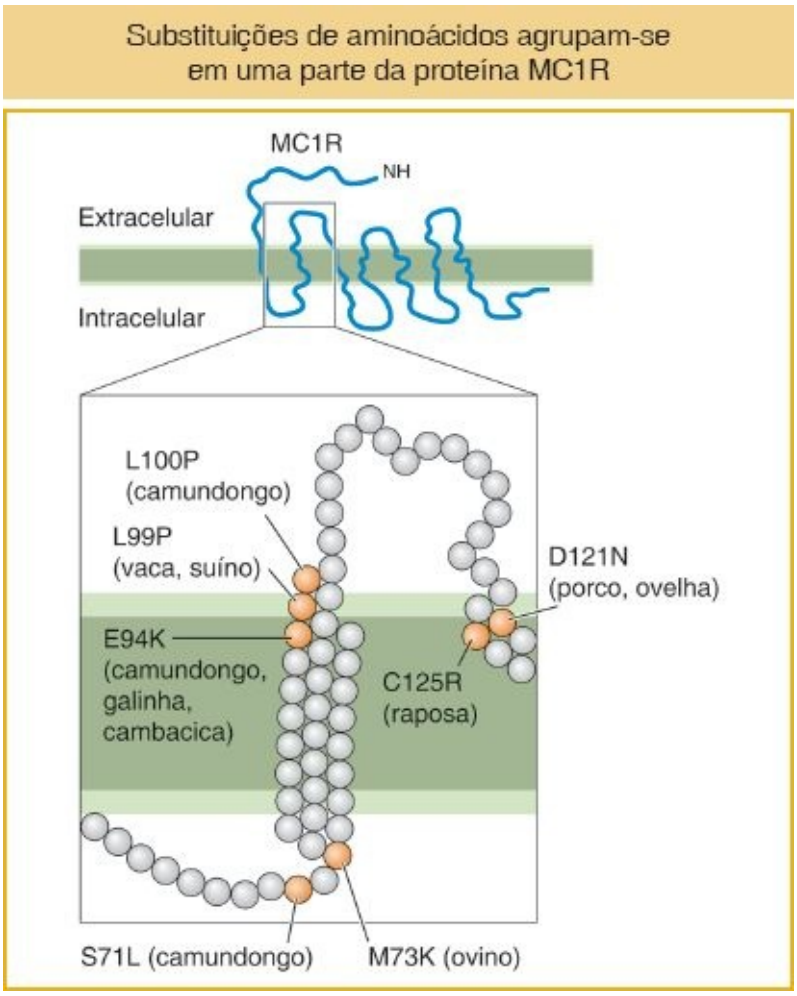


FIGURA 20.12 Substituições de aminoácidos (*círculos laranja*) associadas ao melanismo variam discretamente na localização em diferentes espécies, mas estão localizadas na mesma parte da proteína MC1R. A parte superior da figura demonstra a topologia geral da proteína MC1R. A região na qual as substituições estão localizadas está aumentada na parte inferior da figura. (Dados de E. Eizirik *et al.*, "Molecular Genetics and Evolution in the Cat Family", *Curr. Biol.* 13, 2003, 448-453; reimpressa com permissão de Elsevier.)

Inativação gênica

Há muito tempo observou-se que animais que residem em cavernas com frequência são cegos e incolores. Darwin observou, em *A Origem das Espécies*, que "diversos animais que pertencem às mais diferentes classes, que habitam as cavernas de Carniola (na Eslovênia) e Kentucky (EUA), são cegos. Tendo em vista que é difícil imaginar que os olhos, ainda que inúteis, de algum modo possam ser prejudiciais para os animais que vivem no escuro, a sua perda pode

ser atribuída à não utilização”.[7]

Muitas espécies de peixes que vivem em cavernas perderam seus olhos e a coloração corporal. Tendo em vista que essas espécies pertencem a muitas diferentes famílias que incluem espécies que contêm olhos e que habitam a superfície, claramente a perda dos olhos e da pigmentação ocorreu repetidamente. Por exemplo, o peixe-cego mexicano de caverna (*Astyonax mexicanus*) pertence à mesma ordem que a piranha e o tetra cor de neon. Aproximadamente 30 populações de peixes de caverna no México perderam a coloração do corpo dos seus parentes que habitam a superfície (Figura 20.13).

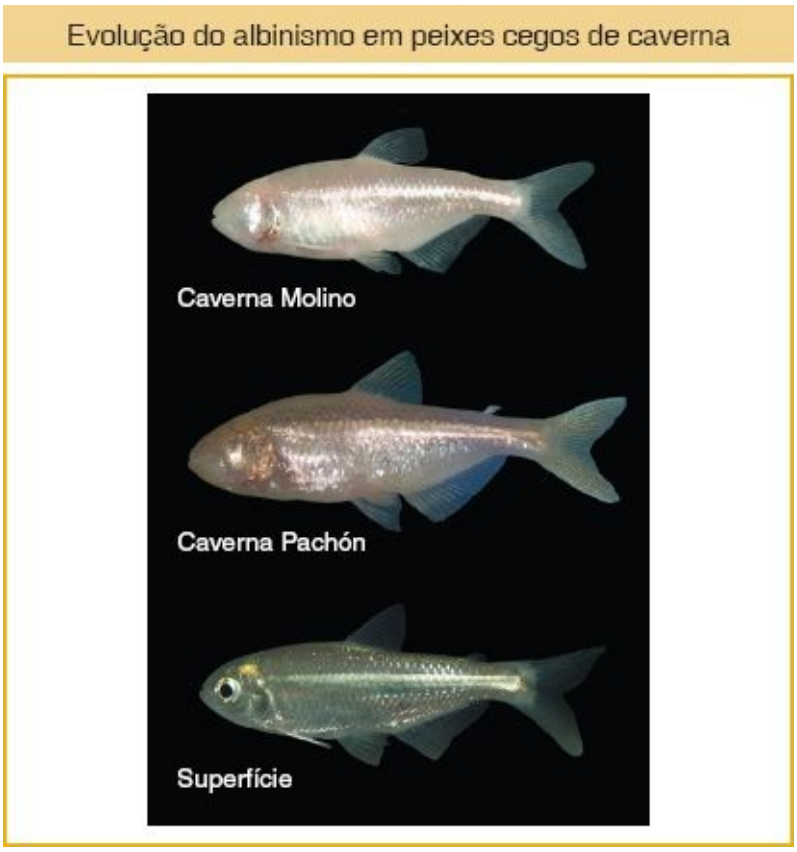


FIGURA 20.13 Tipos de superfície do peixe *Astyanu mexicanus* aparentam ser normais, mas as populações de caverna, tais como aquelas das cavernas de Molino e Pachón, no México, desenvolveram repetidamente a cegueira e o albinismo. (Cortesia de Richard Borowsky.)

Estudos genéticos indicaram que o albinismo na população de peixes da caverna Pachón ocorre em virtude de mutação recessiva única. Além disso, um

cruzamento entre um indivíduo da caverna Molino e um indivíduo da caverna Pachón produziu apenas descendência albina, sugerindo que o albinismo nas duas populações ocorra em virtude do mesmo *locus* genético. Para identificar o gene responsável pelo albinismo nos peixes, pesquisadores investigaram os genótipos de peixes em diversos *loci* de pigmentação que sabidamente causam albinismo em camundongos ou seres humanos. Eles observaram um desses genes, *Oca2*, mapeado no *locus* albino. Eles também observaram que havia uma associação perfeita entre o genótipo do *locus Oca2* e o fenótipo de albinismo na descendência de F_2 que era um retrocruzamento entre a progênie F_1 Molino e Molino/superfície ou a progênie F_1 Pachón e Pachón/superfície.

A inspeção adicional do gene *Oca2* revelou que a população de Pachón era homozigota em relação a uma deleção que se estendia de um íntron até a maior parte de um éxon e que a população Molino era homozigota em relação à deleção de um éxon diferente. Análises funcionais comprovaram que cada deleção no gene *Oca2* causava a perda da função de *Oca2*.

A identificação de diferentes lesões no gene *Oca2* das duas populações de caverna indica que o albinismo evoluiu em separado nas duas populações de caverna. Também existem evidências de que uma terceira população de caverna carrega ainda uma terceira mutação distinta em *Oca2*. Sabe-se de outros vertebrados que o albinismo pode ser desenvolvido por meio de mutações em outros genes. O que pode explicar a repetida inativação do gene *Oca2*? Existem duas explicações possíveis. Primeiramente, as mutações em *Oca2* aparentam não causar nenhum defeito colateral sério além da perda da pigmentação e da visão. Alguns outros genes de pigmentação, quando mutados em peixes, causam reduções dramáticas na viabilidade. Os efeitos das mutações em *Oca2*, então, aparentam ser menos *pleiotrópicos* e causar efeitos sobre a adaptabilidade geral que são menos prejudiciais do que aqueles das mutações em outros genes de pigmentação em peixes. Em segundo lugar, o *locus Oca2* é muito grande, abrangendo aproximadamente 345 kb em seres humanos e contendo 24 éxons. Ele representa um alvo muito grande para mutações aleatórias que perturbariam a função do gene; portanto, é mais provável que surjam mutações em *Oca2* do que mutações em *loci* menores.

A perda da função gênica não é o que normalmente pensamos quando refletimos a respeito da evolução. Mas a inativação gênica certamente é o que devemos prever que ocorra quando as condições seletivas se alteram, ou quando populações ou espécies alteram seus *habitats* ou estilos de vida e determinadas funções gênicas deixam de ser necessárias.

> CONCEITO-CHAVE As mutações de inativação gênica podem ocorrer e aumentar até uma alta frequência quando alterações no *habitat* ou no estilo de vida relaxam a seleção natural sobre traços e funções gênicas subjacentes.

Evolução de sequência reguladora

Conforme discutido anteriormente, uma importante restrição na evolução gênica é o potencial de efeitos colaterais prejudiciais causados por mutações em regiões codificadoras que alteram a função da proteína. Esses efeitos podem ser contornados por mutações nas sequências reguladoras, que são cruciais para a evolução da regulação gênica e para a forma corporal.

Os exemplos de evolução da coloração corporal que vimos até agora apresentam alterações no padrão da pelagem ou das escamas por todo o corpo. A evolução de coloração corporal totalmente preta ou totalmente não pigmentada pode surgir por meio de mutações nos genes de pigmentação. Entretanto, com frequência, muitos esquemas de cores são compostos por duas ou mais cores em algum padrão espacial. Nesses casos, a expressão dos genes de pigmentação deve diferir nas áreas do corpo que serão de cores diferentes. Em diferentes populações ou espécies, a regulação dos genes de pigmentação deve evoluir por meio de algum mecanismo que não altere a função das proteínas de pigmentação.

As espécies das moscas-das-frutas do gênero *Drosophila* demonstram extensiva diversidade de marcas corporais e nas asas. Um padrão comum é a presença de uma mancha preta próxima à ponta das asas em machos (Figura 20.14). A produção de manchas pretas requer enzimas que sintetizam a melanina, o mesmo pigmento produzido em roedores *Chaetodipus intermedius*. Muitos

genes que controlam a via de síntese da melanina foram bem-estudados no organismo-modelo *Drosophila melanogaster*. Um gene é denominado *yellow* (amarelo), tendo em vista que as mutações no gene fazem com que áreas do corpo com pigmentação escura aparentem ser amareladas ou bronze. O gene *yellow* desempenha um papel central no desenvolvimento de padrões de melanina divergentes. Em espécies com manchas, a proteína Yellow é expressa em níveis altos nas células das asas que produzirão a mancha preta, enquanto em espécies sem manchas, a proteína Yellow é expressa em um nível baixo por toda a lâmina das asas (Figura 20.15 A).

FIGURA 20.14 Machos de *Drosophila melanogaster* não apresentam manchas nas asas (*parte superior*), enquanto machos de *Drosophila biarmipes* (*parte inferior*) apresentam manchas escuras nas asas, que são exibidas em um ritual de acasalamento. Esta diferença morfológica simples ocorre em virtude de diferenças na regulação dos genes de pigmentação. (Nicolas Gompel.)

A diferença na expressão de Yellow entre as espécies manchadas e não manchadas pode ocorrer em virtude de diferenças no modo como o gene *yellow* é regulado nas duas espécies. Qualquer um dos dois mecanismos possíveis pode atuar, ou ambos: as espécies podem diferir na disposição espacial de fatores de

transcrição que regulam o gene *yellow* (ou seja, alterações nas sequências de ação *trans* para o gene *yellow*), ou podem diferir nas sequências reguladoras de ação *cis* que controlam como o gene *yellow* é regulado. Para examinar quais mecanismos estão envolvidos, investigadores examinaram a atividade das sequências reguladoras de ação *cis* de *yellow* de diferentes espécies ao posicioná-las *upstream* de um gene repórter e introduzi-las em *D. melanogaster*.

O gene *yellow* é regulado por um arranjo de sequências reguladoras de ação *cis* separadas, que regulam a transcrição do gene em diferentes tecidos e tipos celulares e em diferentes ocasiões no desenvolvimento (Figura 20.15 B). Essas sequências reguladoras incluem aquelas que controlam a transcrição nas partes bucais das larvas, no tórax e no abdome das pupas, e na lâmina das asas em desenvolvimento. Descobriu-se que, enquanto o elemento regulador de ação *cis* da lâmina das asas de espécies não manchadas direciona a expressão em nível baixo de um gene repórter na lâmina das asas, o elemento correspondente de uma espécie manchada, tal como *D. biarmipes* ou *D. elegans*, direciona um alto nível de expressão do repórter em uma mancha próxima da ponta das asas (Figura 20.15 C). Essas observações demonstram que alterações na sequência e na função de um elemento regulador de ação *cis* são responsáveis pela alteração na regulação do gene *yellow* e contribuem para a origem das manchas das asas. Demonstrou-se que o elemento regulador de ação *cis* das espécies manchadas adquiriu sítios de ligação para fatores de transcrição que atualmente direcionam altos níveis de transcrição gênica em um padrão de manchas nas asas em desenvolvimento.

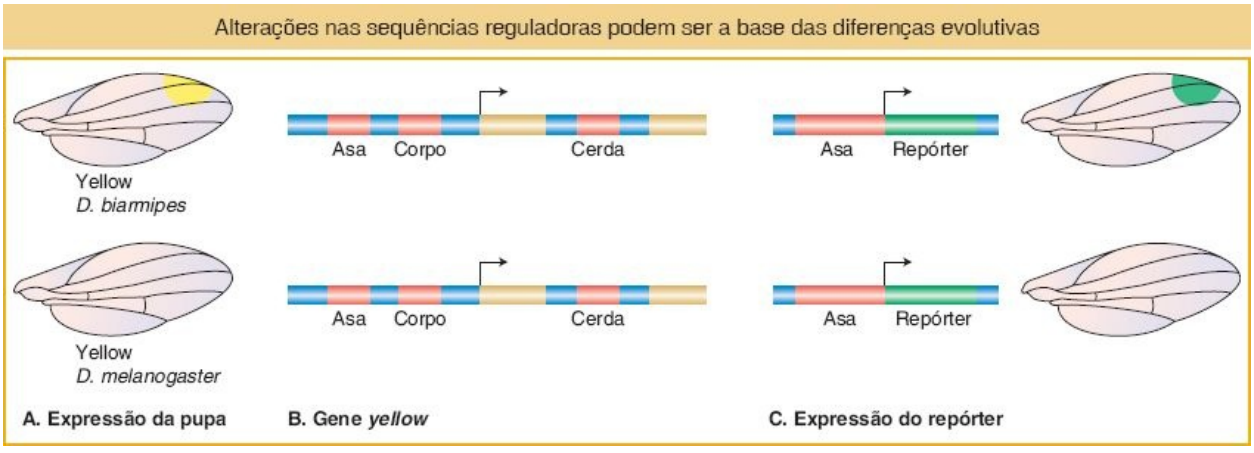


FIGURA 20.15 A evolução da regulação gênica e da morfologia no caso demonstrado ocorre em virtude da evolução em sequências reguladoras de ação *cis*. **A.** Em moscas-das-frutas manchadas, a proteína de pigmentação Yellow é expressa em altos níveis nas células que produzirão grandes quantidades de melanina. **B.** O *locus yellow* das espécies de *Drosophila* contém diversos elementos reguladores de ação *cis* discretos (*vermelho*) que regulam a transcrição de *yellow* em diferentes partes corporais. Os éxons estão demonstrados em dourado. As *setas* indicam o ponto de início e o sentido da transcrição do gene. **C.** O elemento regulador das "asas" da *D. biarmipes* direciona a expressão do gene repórter em um padrão manchado na asa em desenvolvimento, enquanto o elemento homólogo de *D. melanogaster* não ativa um padrão de machas de expressão do repórter. Esta diferença nas atividades do elemento regulador de ação *cis* demonstra que as alterações na função do elemento regulador de ação *cis* são a base das diferenças na expressão da Yellow e na pigmentação entre as duas espécies.

Portanto, alterações evolutivas nas sequências reguladoras de ação *cis* desempenham um papel crítico na evolução da forma corporal. A localização da alteração na sequência reguladora, em vez do próprio gene, pode ser mais bem-explicada à luz dos muitos diferentes efeitos que podem aparecer como resultado de uma mutação codificadora em um gene *toolkit*. Nesse caso, o gene *yellow* é altamente pleiotrópico: ele é necessário para a pigmentação de muitas estruturas e também para funções no sistema nervoso. Uma mutação codificadora que altera a atividade da proteína Yellow alteraria a atividade de Yellow em *todos* os tecidos, o que poderia apresentar uma consequência negativa em termos de aptidão. Entretanto, tendo em vista que as sequências reguladoras de ação *cis* individuais normalmente afetam apenas um aspecto da expressão gênica, mutações nessas sequências proporcionam um mecanismo para alterar um aspecto da expressão gênica, enquanto preservam o papel dos produtos proteicos em outros processos do desenvolvimento.

> CONCEITO-CHAVE As alterações evolutivas nas sequências reguladoras de ação *cis* desempenham um papel crítico na evolução da expressão gênica. Elas contornam os efeitos pleiotrópicos de mutações nas sequências codificadoras dos genes que apresentam múltiplos papéis no desenvolvimento.

Perda de características por meio da evolução de sequências

reguladoras

Características morfológicas podem ser perdidas, bem como ganhas, como resultado de alterações adaptativas nas sequências reguladoras de ação *cis*. Se não houver pressão seletiva para manter uma característica, ela pode ser perdida ao longo do tempo. Mas algumas perdas são benéficas, tendo em vista que facilitam alguma alteração no estilo de vida. Os membros traseiros, por exemplo, foram perdidos muitas vezes em vertebrados – em cobras, lagartos, baleias e peixes-boi – na medida em que esses organismos se adaptaram a diferentes *habitats* e meios de locomoção. As alterações evolutivas em sequências reguladoras de ação *cis* também estão ligadas a essas importantes alterações.

Os precursores evolutivos dos membros traseiros dos vertebrados quadrúpedes são as barbatanas pélvicas dos peixes. Diferenças acentuadas na anatomia das barbatanas pélvicas evoluíram em populações de peixes relacionadas de modo próximo. O peixe esgana-gato ocorre em dois tipos em muitos lagos da América do Norte – um tipo de águas abertas, com espinha pélvica bem-desenvolvida, e um tipo que habita o fundo de águas superficiais, com uma pelve e espinhas dramaticamente reduzidas. Em águas abertas, as espinhas longas ajudam na proteção dos peixes para que não sejam engolidos por predadores maiores. Mas no fundo dos lagos, tais espinhas são uma desvantagem, tendo em vista que podem ser agarradas por larvas de libélulas que se alimentam dos peixes jovens (Figura 20.16 A e 20.16 B).

As diferenças na morfologia pélvica evoluíram repetidamente nos últimos 10.000 anos, desde a recessão das geleiras da última era do gelo. Muitos lagos separados foram colonizados por esgana-gatos oceânicos com espinhas longas, e tipos com espinhas pélvicas reduzidas evoluíram independentemente muitas vezes. Tendo em vista que os peixes estão relacionados de modo tão próximo e são intercruzados em laboratório, geneticistas conseguem mapear os genes envolvidos na redução da pelve. O grupo de David Kingsley, na Stanford University, juntamente com o grupo de Dolph Schluter, na University of British Columbia, mapeou um fator importante envolvido nas diferenças pélvicas no gene *Pitx1*, que codifica um fator de transcrição. Assim como a maior parte dos outros

genes *toolkit* do desenvolvimento, o gene *Pitx1* apresenta diversas funções distintas no desenvolvimento dos peixes. Entretanto, no tipo do peixe esgana-gato com pelve reduzida, a sua expressão é perdida da área do embrião do peixe em desenvolvimento que dará origem ao broto da barbatana pélvica e das espinhas (ver Figura 20.16).

O fato de que a diferença na morfologia pélvica entre os dois tipos foi mapeada no *locus Pitx1* e foi associada à perda da expressão gênica sugeriu que alterações nas sequências reguladoras de *Pitx1* eram responsáveis pela diferença nos fenótipos. Assim como a maior parte dos genes *toolkit* pleiotrópicos, a expressão do gene *Pitx1* em diferentes partes do peixe em desenvolvimento é controlada por elementos reguladores de ação *cis* em separado. Frank Chan *et al.* demonstraram que o elemento regulador que controla a expressão de *Pitx1* na pelve em desenvolvimento foi inativado por grandes mutações de deleção em múltiplas populações independentes de peixes com a pelve reduzida (Figura 20.16 C e D). Além disso, observou-se que a heterozigosidade era reduzida ao redor das sequências de ação *cis* que controlam a expressão pélvica em relação a outras sequências próximas. Essa observação é consistente com a deleção do alelo que está sendo favorecido por seleção natural que atua sobre o tipo com pelve reduzida que habita o fundo.

Portanto, esses achados ilustram ainda mais como mutações em sequências reguladoras contornam os efeitos pleiotrópicos de mutações codificadoras nos genes *toolkit* e que as alterações adaptativas na morfologia podem ocorrer em virtude da perda, bem como do ganho, de expressão gênica durante o desenvolvimento.

> CONCEITO-CHAVE As alterações adaptativas na morfologia podem resultar da inativação de sequências reguladoras e da perda da expressão gênica, bem como da modificação, das sequências reguladoras e do ganho de expressão gênica.

Contornar os efeitos colaterais possivelmente prejudiciais de mutações

codificadoras é um fator muito importante para explicar por que a evolução atua dando origem a novos papéis para os fatores de transcrição que podem regular dúzias a centenas de genes-alvo. Alterações nas sequências codificadoras de um fator de transcrição – por exemplo, no domínio de ligação ao DNA – podem afetar todos os genes-alvo, com consequências catastróficas para o animal. A restrição nas sequências codificadoras de proteínas altamente pleiotrópicas, com muitas funções, explica a extraordinária conservação dos domínios de ligação ao DNA das proteínas Hox e muitos outros fatores de transcrição ao longo de vastos períodos de tempo evolutivo. Embora as funções bioquímicas das proteínas sejam restringidas, a sua regulação diverge. A evolução dos padrões de expressão dos genes *Hox* e outros *toolkit* desempenha um papel importante na evolução da forma corporal.

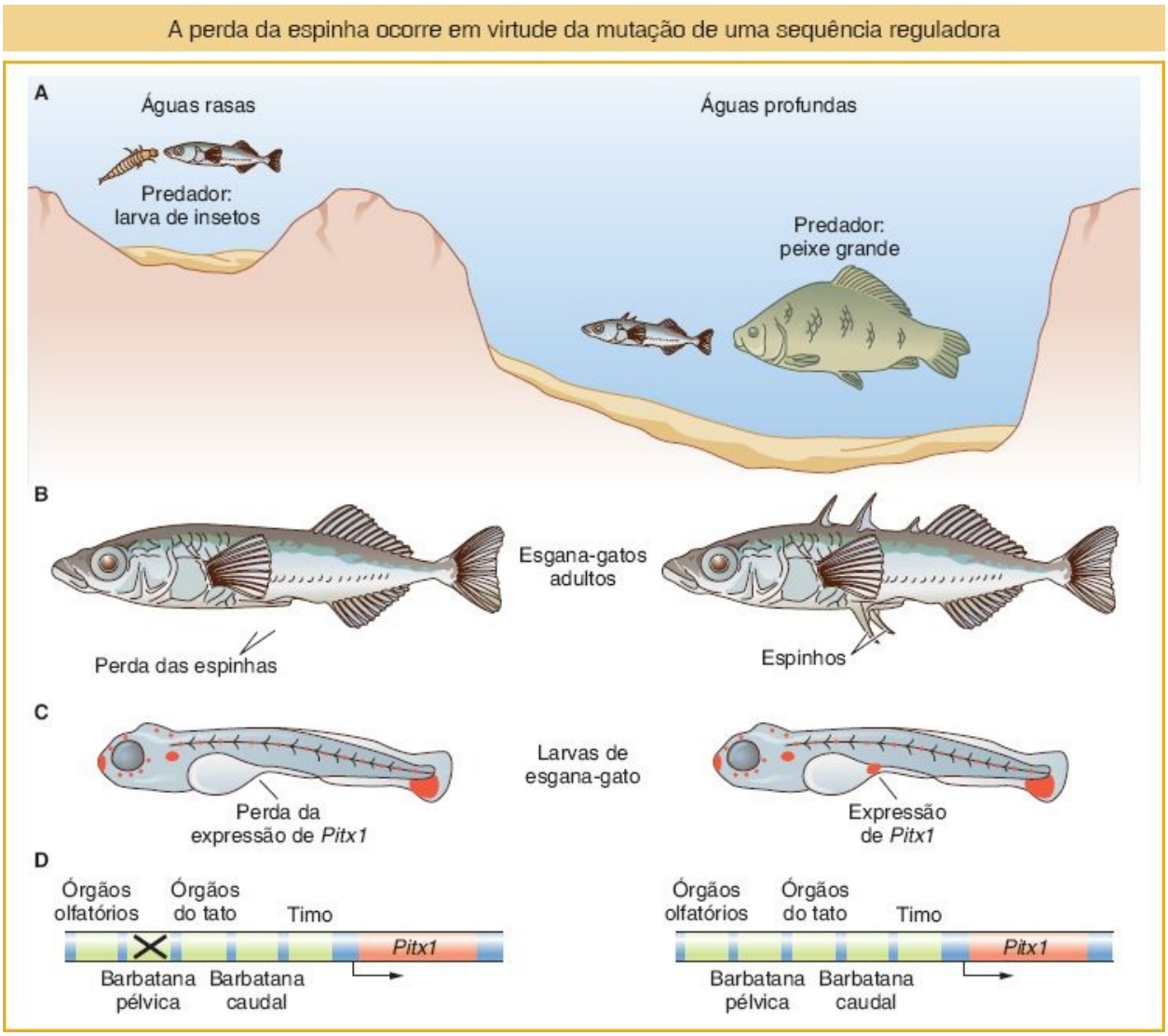


FIGURA 20.16 Deleções em um elemento regulador *cis* de *Pitx1* são a base da evolução adaptativa do esqueleto pélvico do peixe esgana-gato. **A.** Um tipo de peixe esgana-gato habita águas rasas, e um tipo diferente habita águas abertas. **B.** O tipo de águas superficiais apresenta um esqueleto pélvico reduzido (*esquerda*) em relação ao tipo de águas abertas (*direita*). **C.** Esta redução ocorre em virtude da perda seletiva da expressão do gene *Pitx1* (*laranja*) do broto da barbatana pélvica durante o desenvolvimento das larvas do esgana-gato (compare as larvas de esgana-gato à esquerda e à direita). **D.** A perda da expressão de *Pitx1*, por sua vez, ocorre em virtude da mutação de um acentuador do gene *Pitx1* específico para a barbatana pélvica (X marca o acentuador mutado). Outros acentuadores do gene *Pitx1*, que controlam a expressão do gene em algum outro local no corpo em desenvolvimento, não são afetados e atuam de modo semelhante em ambos os tipos de peixe.

Evolução reguladora em seres humanos

A evolução reguladora não está limitada aos genes que afetam o desenvolvimento.

O nível, o momento ou o padrão espacial da expressão de qualquer gene podem variar nas populações ou divergir entre as espécies. Por exemplo, conforme observado anteriormente (ver Capítulo 18), as frequências de alelos no *locus* do grupo sanguíneo *Duffy* variam amplamente em populações humanas. O *locus Duffy* (indicado como *Fy*) codifica uma glicoproteína que atua como um receptor para múltiplas proteínas de sinalização intercelular. Na África Subsaariana, a maior parte dos membros de populações indígenas carrega o alelo Fy^{nulo}. Indivíduos com esse alelo não expressam quaisquer das glicoproteínas Duffy nos eritrócitos, embora a proteína ainda esteja sendo produzida em outros tipos celulares. Como e por que a glicoproteína Duffy está ausente nos eritrócitos desses indivíduos?

A explicação molecular para a ausência de expressão da glicoproteína Duffy nos eritrócitos é a presença de uma mutação de ponto na região promotora do gene *Duffy* na posição –46. Essa mutação está localizada em um sítio de ligação para um fator de transcrição específico de eritrócitos denominado GATA1 (Figura 20.17). A mutação desse sítio elimina a atividade de um acentuador do gene *Duffy* em ensaios de gene repórter.

Uma explicação evolutiva sugere que a ausência da expressão da glicoproteína Duffy nos eritrócitos entre africanos é resultado da seleção natural que favorece a resistência à infecção por malária. O parasita da malária *Plasmodium vivax* é o segundo tipo mais prevalente de parasita da malária na maior parte das regiões tropicais e subtropicais do mundo, mas atualmente está ausente na África Subsaariana. O parasita obtém o acesso aos eritrócitos e aos precursores dos eritrócitos por meio da ligação à glicoproteína Duffy (ver Figura 20.17). A frequência muito alta de homozigotos Fy^{nulo} na África evita que o *P. vivax* se torne comum ali. Além disso, se supusermos que *P. vivax* tenha sido comum na África no passado, o alelo Fy^{nulo} teria então sido selecionado.

A ausência complexa da proteína Duffy nos eritrócitos de uma grande subpopulação dá origem ao questionamento se a proteína Duffy apresenta qualquer função necessária, tendo em vista que aparentemente ela é dispensável. Mas não é o caso de esses indivíduos não apresentarem expressão da proteína Duffy. A proteína é expressa nas células endoteliais do sistema vascular e nas

células de Purkinje do cerebelo. Assim como a evolução da expressão da proteína Yellow em moscas-das-frutas com asas manchadas e da expressão de Pitx1 em esgana-gatos, a mutação reguladora no *locus Fy* possibilita que um aspecto da expressão gênica (em eritrócitos) seja alterado sem modificar os outros (ver Figura 20.17).

Modificações em sequências codificadoras e reguladoras são meios comuns para a alteração evolutiva. Elas ilustram como a diversidade por surgir sem a alteração no número de genes em uma espécie. Entretanto, as alterações mutacionais em maior escala podem ocorrer no DNA, e ocorrem, o que resulta na expansão do número de genes, e essa expansão fornece a matéria-prima para a inovação evolutiva.

20.6 Origem de novos genes e funções proteicas

A evolução é composta por mais do que a substituição de um alelo por outro em *loci* de funções definidas. Uma grande fração de genes codificadores de proteínas e de RNA pertence a **famílias gênicas**, grupos de genes que estão relacionados em termos de sequência e tipicamente também na função bioquímica. Por exemplo, existem mais de 1.000 genes que codificam receptores olfatórios estruturalmente relacionados em um camundongo e três genes de opsina estruturalmente relacionados que codificam proteínas necessárias para a visão em cores nos seres humanos. Dentro de famílias como essas, foram desenvolvidas novas funções, que possibilitaram novas capacidades. Essas novas funções podem ser expansões de capacidades existentes. Nos exemplos anteriores, novos receptores apareceram em camundongos com a capacidade de detectar novas substâncias químicas no ambiente ou, no caso dos seres humanos e de seus parentes primatas do Velho Mundo, apareceram novas opsinas que conseguem detectar comprimentos de onda de luz que outros mamíferos não conseguem. Em outros casos, a evolução de novas famílias gênicas pode levar a funções totalmente novas, que podem proporcionar novos modos de vida, tais como a aquisição de proteínas anticongelamento em peixes polares. Aqui, indagaremos,

de onde vem o DNA para os novos genes? Quais são os destinos dos novos genes? E como as novas funções proteicas se desenvolvem?

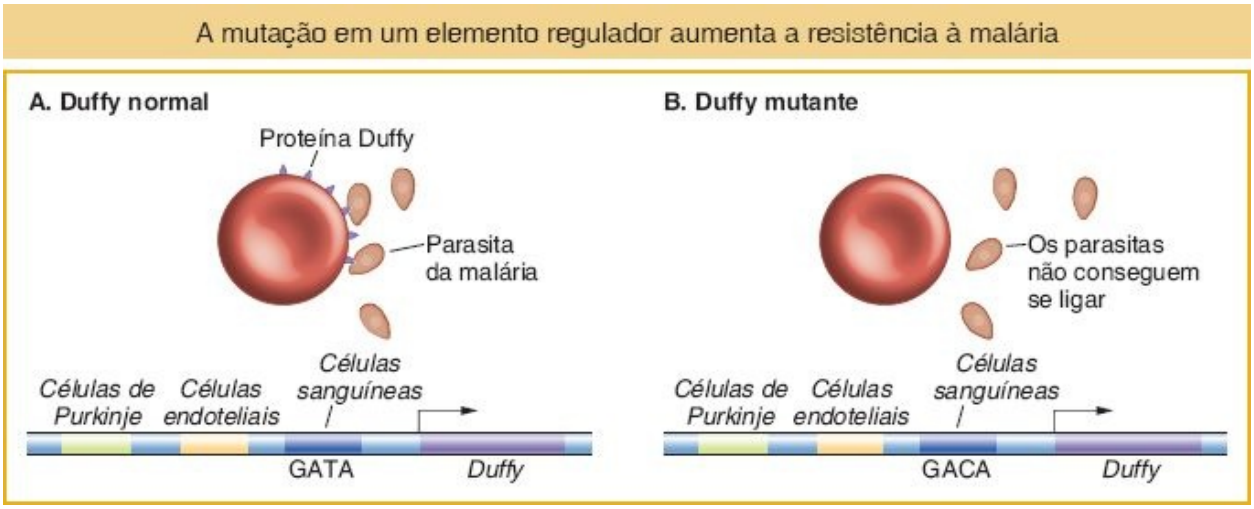


FIGURA 20.17 Uma mutação reguladora em um acentuador do gene humano *Duffy* está associada à resistência à malária. **A.** A proteína Duffy (*azul-escuro*) é tipicamente expressa em células sanguíneas, bem como em células de Purkinje no cérebro e em células endoteliais. **B.** Uma alta proporção de africanos ocidentais não apresenta expressão de Duffy em seus eritrócitos em virtude de uma mutação em um acentuador de células sanguíneas (a sequência GATA é mutada para GACA). Tendo em vista que a proteína Duffy é parte do receptor para o parasita da malária *P. vivax* (*laranja*), indivíduos com a mutação reguladora são resistentes à infecção, mas apresentam expressão normal de Duffy em algum outro local no corpo.

Expansão do número de genes

Existem diversos mecanismos genéticos que podem expandir o número de genes ou de partes de genes. Um processo em grande escala para a expansão do número de genes é a formação de *poliploides*, indivíduos com mais de dois conjuntos cromossômicos. Os poliploides resultam da duplicação do genoma inteiro. Muito mais comum em plantas do que em animais (ver Capítulo 17), a formação de poliploides desempenhou um papel importante na evolução das espécies de plantas. Considere a distribuição de frequências de números de cromossomos haploides entre as espécies de plantas dicotiledôneas demonstradas na Figura 20.18. Acima de um número cromossômico de aproximadamente 12, números pares são muito mais comuns do que números ímpares – uma consequência da poliploidia frequente.

Um segundo mecanismo que pode aumentar o número de genes é a **duplicação gênica**. A replicação errônea do DNA durante a meiose pode causar a duplicação de segmentos de DNA. Os comprimentos dos segmentos duplicados podem variar de apenas um ou dois nucleotídios até segmentos substanciais de cromossomos, contendo muitos ou até mesmo centenas de genes. Análises detalhadas da variação do genoma humano revelaram que seres humanos comumente carregam pequenas duplicações que resultam em uma variação no número de cópias de genes.

Um terceiro mecanismo que pode dar origem às duplicações gênicas é a *transposição*. Por vezes, quando um elemento de transposição é transposto para outra parte do genoma, ele pode carregar também material genético adicional do hospedeiro e inserir uma cópia de alguma parte do genoma em outro local (ver Capítulo 15).

Um quarto mecanismo que pode expandir o número de genes é a **retrotransposição**. Muitos genomas animais abrigam elementos genéticos semelhantes a retrovírus (ver Capítulo 15), que codificam atividade de transcriptase reversa. Os próprios retrotranspósons compõem aproximadamente 40% do genoma humano. Ocasionalmente, transcritos de mRNA do genoma hospedeiro são transcritos de modo reverso em cDNA e inseridos de volta no genoma, produzindo uma duplicata do gene sem íntrons.

Destino de genes duplicados

Chegou-se a acreditar que, em virtude de a função ancestral ser proporcionada pelo gene original, os genes duplicados seriam essencialmente elementos genéticos sobressalentes, que estão livres para desenvolver novas funções (denominada **neofuncionalização**) e que esse seria um destino comum. Entretanto, a análise detalhada de genomas e as considerações da genética de populações levaram a melhor compreensão sobre os destinos alternativos das novas duplicatas de genes, com a evolução da nova função sendo apenas uma via.

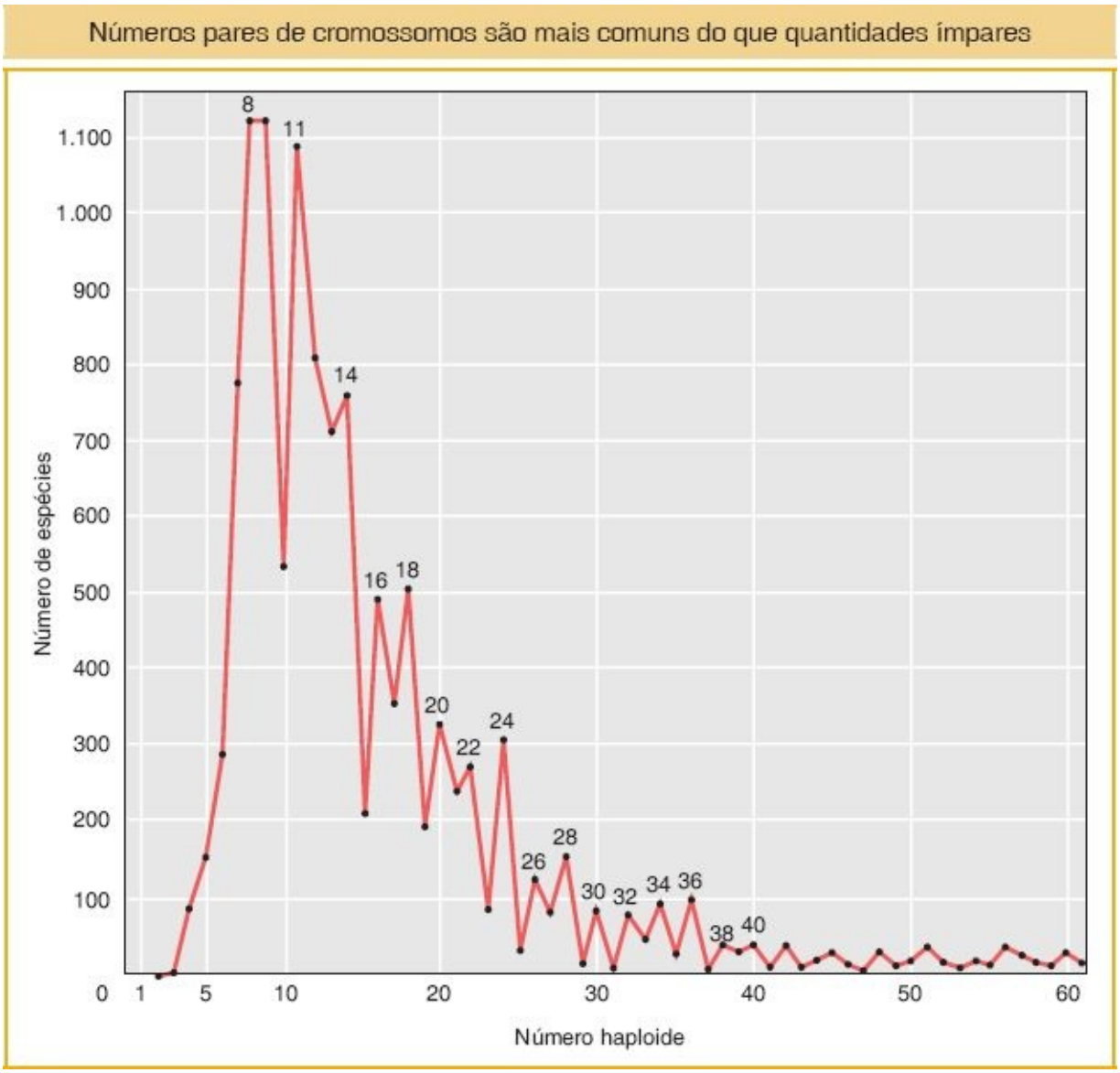


FIGURA 20.18 Distribuição de frequência de números haploides de cromossomos em plantas dicotiledôneas. (Dados de Verne Grant, *The Origin of Adaptations*. Columbia University Press, 1963.)

Para simplificar, consideraremos um evento de duplicação que resulta na duplicação de toda a região codificadora e reguladora de um gene (Figura 20.19 A). Muitos desfechos diferentes podem surgir a partir de uma duplicação. O resultado mais simples é que o alelo que contém a duplicação é perdido da população antes de aumentar até alcançar qualquer frequência significativa, assim como é o destino de muitas novas mutações (ver Capítulo 18). Mas consideraremos em seguida cenários mais interessantes: suponha que a

duplicação sobreviva e que novas mutações comecem a ocorrer no par de genes duplicado. Tendo em mente que os genes originais e duplicados são inicialmente cópias exatas e, portanto, redundantes, após o surgimento de novas mutações, existem diversos destinos possíveis:

1. Pode ocorrer uma mutação de inativação na região codificadora de qualquer duplicata. O parálogo inativado é denominado **pseudogene** e em geral será invisível para a seleção natural. Portanto, ele acumulará mais mutações e evoluirá por deriva genética, enquanto a seleção natural manterá o parálogo funcional (Figura 20.19 B).
2. Podem ocorrer mutações que alteram a regulação de uma duplicata ou a atividade de uma proteína codificada. Esses alelos podem então se tornar sujeitos à seleção positiva e adquirir uma nova função (neofuncionalização) (Figura 20.19 C).
3. Em casos nos quais o gene ancestral apresenta mais de uma função e mais de um gene regulador, assim como em relação à maior parte dos genes *toolkit*, um terceiro desfecho possível é que as mutações iniciais inativem ou alterem um elemento regulador em cada duplicata. A função do gene inicial agora é dividida entre as duplicatas, que se complementam. Com a finalidade de preservar a função ancestral, a seleção natural manterá a integridade de ambas as regiões codificadoras dos genes. Diz-se que os *loci* que seguem essa via de duplicação e mutação que produzem parálogos complementares são **subfuncionalizados** (Figura 20.19 D).

Alguns desses destinos alternativos de duplicatas de genes estão ilustrados na história da evolução gênica da globina humana. A evolução da nossa linhagem, desde os ancestrais peixes até o amniotos terrestres que botam ovos e até os mamíferos placentários, necessitou de uma série de inovações na oxigenação tecidual, que incluem a evolução de genes de globina adicionais, com novos padrões de regulação, e a evolução de proteínas de hemoglobina com propriedades de ligação ao oxigênio distintas.

A hemoglobina adulta é um tetrâmero composto por duas cadeias de polipeptídios α e duas cadeias β, cada uma com a sua molécula de heme ligada. O

gene que codifica a cadeia α adulta está no cromossomo 16 e o gene que codifica a cadeia β está no cromossomo 11. As duas cadeias são aproximadamente 49% idênticas em suas sequências de aminoácidos; essa similaridade reflete a sua origem comum a partir de um gene de globina ancestral há muito no tempo evolutivo. O gene da cadeia α está localizado em um aglomerado de cinco genes relacionados (α e ζ) no cromossomo 16, enquanto o da cadeia β está localizado em um aglomerado de seis genes relacionados no cromossomo 11 (ε, β, δ e γ) (Figura 20.20). Cada aglomerado contém um pseudogene, ψ_α e ψ_β, respectivamente, que acumulou mutações inativadoras aleatórias.

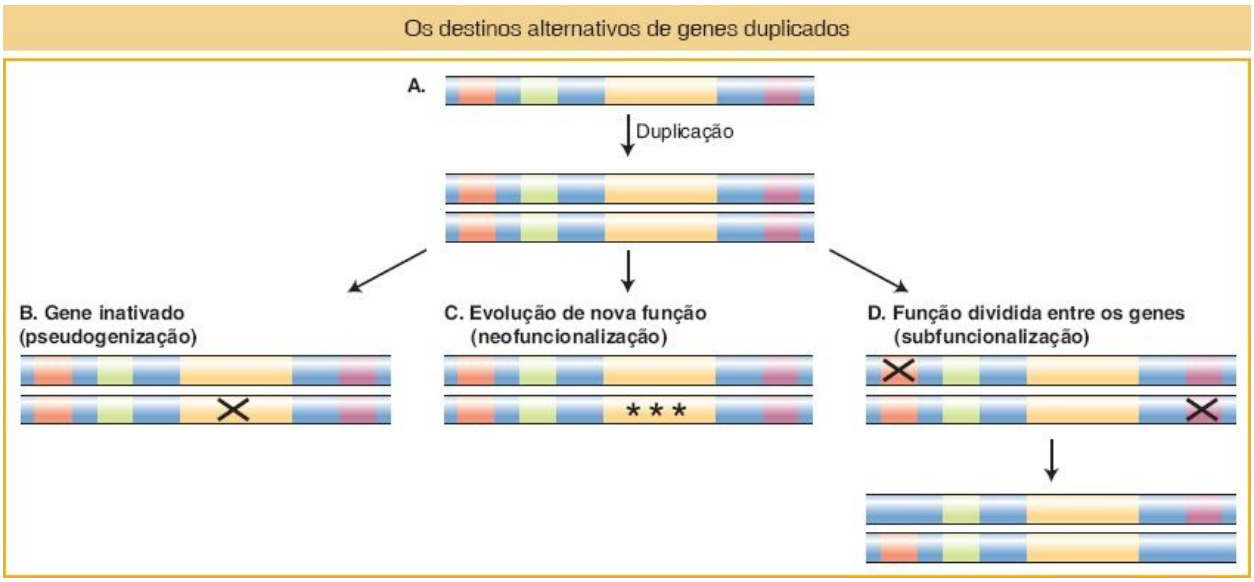


FIGURA 20.19 Os destinos alternativos de genes duplicados. **A.** A duplicação de um gene. Os quadros *laranja*, *amarelos* e *cor-de-rosa* indicam elementos reguladores *cis*; o quadro *bege* indica a região codificadora. Após a duplicação, diversos destinos alternativos das duplicatas são possíveis: (**B**) qualquer mutação inativadora em uma região codificadora tornará aquela duplicata um pseudogene e, em seguida, a seleção purificadora operará no parálogo remanescente; (**C**) podem surgir mutações que alteram a função de uma proteína, e elas podem ser favorecidas por seleção positiva (neofuncionalização); (**D**) as mutações podem afetar uma subfunção de qualquer duplicata, e, desde que os dois parálogos em conjunto proporcionem as funções ancestrais, diferentes subfunções podem ser mantidas, resultando na evolução de dois *loci* complementares (subfuncionalização).

Cada aglomerado contém genes que desenvolveram perfis de expressão distintos, uma função distinta, ou ambos. De maior interesse são os dois genes γ. Esses genes são expressos durante os últimos 7 meses de desenvolvimento fetal

para produzir a hemoglobina fetal (também conhecida como hemoglobina F), que é composta por duas cadeias α e duas cadeias γ. A hemoglobina fetal apresenta maior afinidade pelo oxigênio do que a hemoglobina adulta, o que possibilita que o feto extraia o oxigênio a partir da circulação materna por meio da placenta. Ao nascimento, até 95% da hemoglobina é do tipo fetal; em seguida, a expressão da forma β adulta substitui a γ e uma pequena quantidade de globina δ também é produzida. A ordem de aparecimento das cadeias de globina durante o desenvolvimento é orquestrada por um conjunto complexo de sequências reguladoras de ação *cis* e, notavelmente, segue a ordem dos genes em cada cromossomo.

Os genes γ são restritos aos mamíferos placentários. Sua regulação do desenvolvimento distinta e seus produtos proteicos significam que essas duplicatas desenvolveram diferenças na função que contribuíram para a evolução do estilo de vida placentário. Curiosamente, são conhecidas variantes reguladoras desses genes, que causam a persistência da expressão da hemoglobina fetal na infância e na fase adulta. Essas variantes de ocorrência natural aparentam moderar a gravidade da anemia falciforme ao suprimir os níveis de Hb^S produzidos. Uma estratégia difundida para o tratamento da anemia falciforme é administrar fármacos que estimulam a reativação da expressão da hemoglobina fetal.

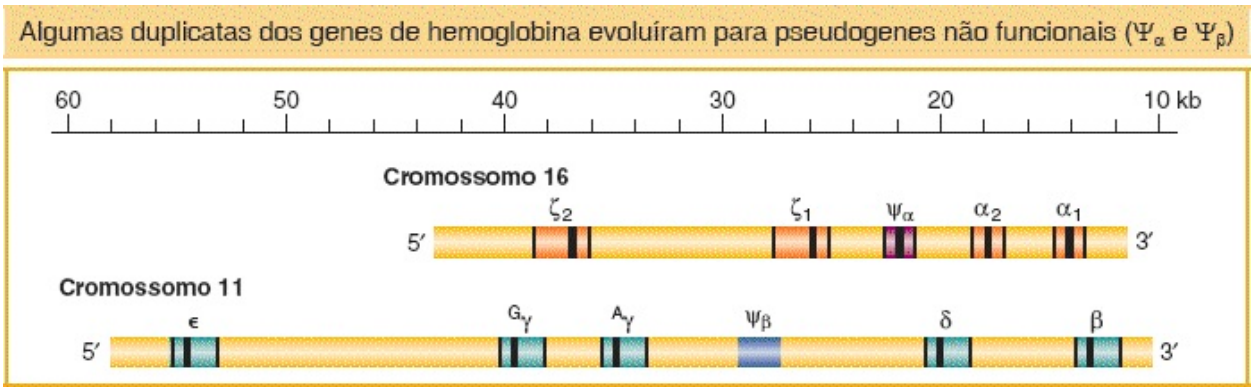


FIGURA 20.20 Distribuição cromossômica dos genes em relação à família α de globinas no cromossomo 16 e à família β de globinas no cromossomo 11 em seres humanos. A estrutura do gene está demonstrada pelas *barras pretas* (éxons) e pelas *barras coloridas* (íntrons).

RESUMO

A teoria da evolução por seleção natural explica as alterações que ocorrem em populações de organismos como resultantes de alterações nas frequências relativas de diferentes variantes na população. Se não houver variação dentro de uma espécie em relação a algum traço, não pode haver evolução. Além disso, tal variação deve ser influenciada por diferenças genéticas. Se as diferenças não forem hereditárias, elas não podem evoluir, tendo em vista que a vantagem reprodutiva de uma variante não será transmitida para as gerações seguintes. É crucial compreender que os processos mutacionais que dão origem à variação no genoma atuam aleatoriamente, mas que o processo seletivo que separa as variantes vantajosas e desvantajosas não é aleatório.

A capacidade de estudar a evolução no nível do DNA e das proteínas transformou a nossa compreensão sobre o processo evolutivo. Antes que tivéssemos a capacidade de estudar a evolução no nível molecular, não havia suspeita de que grande parte da evolução fosse, de verdades, resultado de deriva genética, não da seleção natural. Uma grande parte da evolução molecular aparenta ser a substituição de uma sequência proteica por outra de função equivalente. Entre as evidências em relação à prevalência da evolução neutra está o fato de que o número de diferenças nos aminoácidos entre duas espécies diferentes em uma molécula – por exemplo, hemoglobina – é diretamente proporcional ao número de gerações desde a sua divergência a partir de um ancestral comum no passado evolutivo. Não esperaríamos que existisse um "relógio molecular" com uma taxa de alterações constante se a seleção das diferenças fosse dependente de alterações em particular no ambiente.

Assim, uma grande parte da evolução de sequências é neutra, de modo que não há uma relação simples entre a quantidade de alterações na sequência de DNA de um gene e a quantidade de alterações, se existente, na função da proteína codificada. Algumas funções proteicas podem ser alteradas por meio da substituição de um único aminoácido, enquanto outras requerem uma série de substituições ocasionadas por meio da seleção cumulativa. Tais vias adaptativas de múltiplas etapas podem seguir vias diferentes, inclusive quando as condições

de seleção natural são as mesmas. Isso ocorre porque as vias disponíveis para qualquer população em determinado momento dependem da ocorrência de mutações ao acaso, que podem não surgir na mesma ordem em diferentes populações. Além disso, as etapas anteriores adotadas podem afetar o fato de uma nova mutação ser favorecida, desfavorecida ou neutra.

Antes do advento da genética molecular, não era possível saber se eventos evolutivos independentes poderiam ter dado origem à mesma adaptação múltiplas vezes. Ao detalhar os genes e as mutações exatas envolvidas nas alterações de função, agora reconhecemos que a evolução pode se repetir, e se repete, por meio da ação sobre os mesmos genes para produzir resultados semelhantes em casos independentes. Por exemplo, alterações nos mesmos genes são responsáveis por casos de surgimento independente de melanismo e albinismo em alguns vertebrados, ou pela perda de espinhas pélvicas em diferentes populações de peixes esgana-gato. A evolução pode se repetir por meio da alteração do mesmo nucleotídio no caso do surgimento independentemente das mutações falciformes que levam à resistência adaptativa à malária.

Uma importante restrição na evolução das sequências codificadoras é a dos efeitos colaterais possivelmente prejudiciais das mutações. Se uma proteína desempenha múltiplas funções em diferentes tecidos, como é o caso em relação a muitos genes envolvidos na regulação dos processos de desenvolvimento, mutações nas sequências codificadoras podem afetar todas as funções e diminuir a adaptabilidade. Os possíveis efeitos pleiotrópicos das mutações codificadoras podem ser contornados por mutações em sequências reguladoras não codificadoras. As mutações nessas sequências podem alterar seletivamente a expressão gênica em apenas um tecido ou uma parte corporal, não em outras. A evolução das sequências reguladoras de ação *cis* é central para a evolução dos traços morfológicos e para a expressão de genes *toolkit* que controlam o desenvolvimento.

Com frequência surgem novas funções proteicas por meio da duplicação de genes e subsequente mutação. Um novo DNA pode surgir por meio da duplicação do genoma inteiro (poliploidia), uma ocorrência frequente em plantas, ou por meio de diversos mecanismos que produzem duplicatas de genes individuais ou

conjuntos de genes. O destino de genes duplicados depende muito da natureza das mutações adquiridas após a duplicação. Os possíveis destinos são a inativação de uma duplicata, a divisão de função entre duas duplicatas, ou o ganho de novas funções.

Em geral, a evolução genética está sujeita à contingência histórica e ao acaso, mas é restringida pela necessidade de os organismos sobreviverem e se reproduzirem em um mundo constantemente em alteração. O "mais adaptado" é um estado condicional, sujeito a alteração na medida em que o planeta e os *habitats* são alterados.

TERMOS-CHAVE

adaptação
duplicação gênica
epistasia de sinal
família gênica
neofuncionalização
pseudogene
relógio molecular
retrotransposição
seleção cumulativa
seleção natural
subfuncionalização
substituição não sinônima
substituição sinônima
via adaptativa

PROBLEMAS RESOLVIDOS

Problema resolvido 1. Observa-se que duas espécies de bactérias relacionadas de modo próximo fixaram dois alelos diferentes detectados eletroforeticamente em um *locus* que codifica uma enzima envolvida na fragmentação de um nutriente. Como você testaria experimentalmente se a divergência nas sequências

enzimáticas pode ter causado diferenças na função e na adaptabilidade?

Solução

Com a finalidade de testar se as enzimas apresentam diferentes propriedades funcionais, devemos planejar experimentos *in vitro* e *in vivo*. Se os substratos e as propriedades das enzimas forem conhecidos, podemos purificar a enzima de cada espécie e medir diretamente se existem diferenças funcionais. Alternativamente, um teste indireto seria verificar se cada espécie cresce bem no nutriente em particular que a enzima fragmenta.

Idealmente, com a finalidade de medir as diferenças na adaptabilidade, substituímos a região codificadora da enzima de uma espécie pela região codificadora da enzima da segunda espécie, e vice-versa. Em seguida, o crescimento de cada linhagem do tipo selvagem e transgênica poderia ser comparado no meio contendo o mesmo nutriente, com o crescimento sendo um indicador de adaptabiliadade. Se houver diferenças na adaptabilidade relativa das linhagens transgênica e do tipo selvagem, então é possível que as enzimas tenham divergido sob a seleção natural. Caso negativo, é então provável que as enzimas tenham evoluído de modo neutro ou que o efeito da seleção seja muito pequeno para ser medido experimentalmente.

PROBLEMAS

QUESTÕES SOBRE AS FIGURAS

1. Na Figura 20.5, observe que a diferença nas taxas de sobrevida entre os genótipos *AS* e *AA* declina na medida em que as crianças ficam mais velhas. Ofereça uma possível explicação para essa observação.
2. Examinando a Figura 20.8, explique por que a taxa de evolução em sítios não sinônimos é mais baixa. Você espera que isso seja verdadeiro apenas para os genes de globina ou para a maior parte dos genes?
3. A partir da Tabela 20.3, você espera que a mutação não codificadora g4205a seja fixada antes ou depois da mutação codificadora G238S em

uma população de bactérias que está desenvolvendo resistência ao antibiótico cefotaxima? Forneça no mínimo dois motivos para a sua resposta.

4. Examinando a Tabela 20.4, qual você acredita que seja a ordem das mutações fixadas durante a seleção em uma terceira linhagem de vírus em evolução? As mutações se tornariam fixadas na mesma ordem que o vírus TX ou ID?
5. Examinando a Tabela 20.5, como a interpretação dos resultados do teste de McDonald-Kreitman diferiria se o número de diferenças de espécies observadas não sinônimas fosse 1 em vez de 7?
6. Utilizando a Figura 20.17, explique como a mutação na sequência GATA do gene *Duffy* confere resistência à infecção pelo *P. vivax*.
7. Na Figura 20.18, qual é a evidência de que a formação de poliploides foi importante na evolução das plantas?

PROBLEMAS BÁSICOS

8. Compare a descrição de Darwin sobre a seleção natural, conforme citado anteriormente neste capítulo, com a descrição de Wallace sobre a tendência das variedades de se distanciarem do tipo original descrita logo após. Que ideias elas apresentam em comum?
9. Quais são os três princípios da teoria da evolução por seleção natural?
10. Por que a teoria neutra da evolução molecular foi uma ideia revolucionária?
11. Qual você preveria que fosse a taxa relativa de substituições sinônimas e não sinônimas em um pseudogene de globina?
12. Os heterozigotos *AS* são completamente resistentes à infecção por malária? Explique a evidência para a sua resposta.

PROBLEMAS DESAFIADORES

13. Se a taxa de mutação de um novo alelo é 10^{-5}, presumindo que não existe migração, quão grandes as populações isoladas devem ser para prevenir a diferenciação ao acaso entre elas na frequência desse alelo?

14. A glicose-6-fosfato desidrogenase (G6PD) é uma enzima crítica envolvida no metabolismo da glicose, especialmente em eritrócitos. Deficiências na enzima são o defeito enzimático humano mais comum e ocorrem a uma alta frequência em determinadas populações de crianças da África Oriental.

a. Ofereça uma hipótese para a alta incidência de mutações na G6PD em crianças da África Oriental.

b. Como você testaria adicionalmente a sua hipótese?

c. Pontos de diferentes mutações na G6PD que afetam a função enzimática foram encontrados em populações humanas. Ofereça uma explicação para a abundância de diferentes mutações na G6PD.

15. Grandes diferenças nas frequências de Hb^S entre tribos quenianas e ugandenses têm sido observadas em análises conduzidas por outros pesquisadores além de Tony Allison. Esses pesquisadores ofereceram explicações alternativas, diferentes da ligação com a malária proposta por Allison. Ofereça um contra-argumento, ou um teste experimental em relação às seguintes hipóteses alternativas:

a. A taxa de mutação é mais alta em determinadas tribos.

b. Existe um baixo grau de mistura genética entre as tribos e, assim, o alelo aumentou sua frequência por meio do endocruzamento em determinadas tribos.

16. Quantas possíveis vias evolutivas existem para que um alelo desenvolva seis mutações diferentes? Sete mutações diferentes? Dez mutações diferentes?

17. O gene *MC1R* afeta a cor da pele e dos cabelos em seres humanos. Existem no mínimo 13 polimorfismos do gene nas populações europeias e asiáticas, 10 dos quais são não sinônimos. Em africanos, existem no mínimo 5 polimorfismos do gene, nenhum dos quais é não sinônimo. Qual pode ser uma explicação para as diferenças na variação do *MC1R* entre africanos e

não africanos?

18. As proteínas opsinas detectam a luz em células fotorreceptoras do olho e são necessárias para a visão em cores. O macaco-da-noite, o lêmure noturno e roedores da família Spalacidae (semelhantes à toupeira) apresentam diferentes mutações em um gene da opsina, que a torna não funcional. Explique por que todas as três espécies conseguem tolerar mutações nesse gene, que opera na maior parte dos outros mamíferos.

19. A ausência total ou parcial de membros evoluiu muitas vezes em vertebrados (cobras, lagartos, peixes-boi, baleias). Você espera que as mutações que ocorreram na evolução da ausência de membros estejam nas sequências codificadoras ou não codificadoras de genes *toolkit*? Por quê?

20. Diversas espécies de *Drosophila* com asas não manchadas são descendentes de um ancestral manchado. Você preveria que a perda da formação de manchas envolvesse alterações codificadoras ou não codificadoras nos genes de pigmentação? Como você testaria qual é o caso?

21. Foi declarado que a "evolução se repete". Qual é a evidência para essa declaração a partir:

a. Da análise dos alelos Hb^S?

b. Da análise da resistência a antibióticos em bactérias?

c. Da análise do bacteriófago ΦX174 experimentalmente selecionado?

d. Da análise de mutações no *Oca2* em peixes de caverna?

e. Da análise dos *loci Pitx1* do esgana-gato?

22. Qual é a evidência molecular de que a seleção natural inclui a "rejeição de alterações prejudiciais"?

23. Quais são os três destinos alternativos de uma nova duplicata de gene?

24. Qual é a evidência de que a duplicação gênica tem sido a fonte das famílias de genes α e β em relação à hemoglobina humana?

25. Estudos de sequenciamento do DNA em relação a um gene em duas

espécies relacionadas de modo próximo produzem os seguintes números de sítios que variam:

Polimorfismos sinônimos	50
Polimorfismos não sinônimos	20
Diferenças entre espécies sinônimas	18
Diferenças entre espécies não sinônimas	2

Esse resultado ampara a evolução neutra do gene? Ele ampara uma substituição adaptativa de aminoácidos? Qual explicação você ofereceria para as observações?

26. Em seres humanos, observa-se que dois genes que codificam os pigmentos visuais da opsina que são sensíveis aos comprimentos de onda de luz verde e vermelha estão adjacentes entre si no cromossomo X. Eles codificam proteínas que são 96% idênticas. Mamíferos não primatas possuem apenas um gene que codifica uma opsina sensível ao comprimento de onda vermelho/verde.

a. Ofereça uma explicação para a presença dos dois genes de opsina no cromossomo X humano.

b. Como você testaria sua explicação e detalharia quando na história evolutiva o segundo gene surgiu?

27. Aproximadamente 9% dos homens caucasianos são daltônicos e não conseguem distinguir os objetos de cor vermelha dos objetos de cor verde.

a. Ofereça um modelo genético em relação ao daltonismo.

b. Explique por que e como o daltonismo alcançou uma frequência de 9% nessa população.

[1]C. Darwin, *Charles Darwin's Beagle Diary*, Ed. R. D. Keynes, Cambridge University Press, 2001.

[2]C. Darwin, *Charles Darwin's Beagle Diary*, Ed. R. D. Keynes, Cambridge University Press, 2001.

[3]C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life*, D. Appleton, 1864.

[4]C. Darwin e A. Wallace, "On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection", *Journal of the Proceedings of the Linnean Society of London. Zoology* 3, 1858, 45-50.

[5]A. C. Allison, "Protection Afforded by Sickle-cell Trait against Subtertian Malarial Infection", *British Medical Journal* 1, 1954, 290-294.

[6]G. G. Simpson, "Organisms and Molecules in Evolution", *Science* 146, 1964, 1535-1538.

[7]C. Darwin, *On the Origin of Species by Means of Natural Selection*, p. 137. John Murray, Londres, 1859.

Guia Resumido de Organismos-modelo

Escherichia coli • *Saccharomyces cerevisiae* • *Neurospora crassa* • *Arabidopsis thaliana* • *Caenorhabditis elegans* • *Drosophila melanogaster* • *Mus musculus*

Este guia resumido reúne em um local as principais características dos organismos-modelo relacionados à genética. Cada um dos sete organismos-modelo recebe sua própria extensão de duas páginas; o formato é consistente, possibilitando que os leitores comparem e contrastem as características dos organismos-modelo. Cada tratamento enfoca características especiais do organismo que o tornaram útil como um modelo; as técnicas especiais que foram desenvolvidas para o estudo do organismo; e as principais contribuições que os estudos do organismo realizaram para a nossa compreensão sobre a genética. Embora muitas diferenças estejam aparentes, as abordagens das análises de genética em geral são semelhantes, mas precisam ser adaptadas para levar em consideração o ciclo de vida individual, o nível de ploidia, tamanho e formato, e as propriedades genômicas, tais como a presença de plasmídios e transpósons naturais.

Os organismos-modelo sempre estiveram na vanguarda da genética. Inicialmente, no desenvolvimento histórico de um organismo-modelo, um pesquisador seleciona o organismo em virtude de alguma característica que o torna particularmente apropriado para o estudo de um processo genético no qual o pesquisador está interessado. O conselho dos últimos cem anos foi: “escolha bem o seu organismo.” Por exemplo, os fungos ascomicetos, tais como *Saccharomyces cerevisiae* e *Neurospora crassa*, são muito apropriados para o estudo dos processos meióticos, tais como o *crossing over*, em virtude de sua característica única, o asco, manter unidos os produtos de uma única meiose.

Diferentes espécies tendem a demonstrar processos notavelmente semelhantes, até mesmo entre os membros de grandes grupos, tais como os eucariotos. Portanto, podemos esperar razoavelmente que o que se aprende em uma espécie

pode ser, no mínimo parcialmente, aplicado para outras. Em particular, os geneticistas mantêm os olhos abertos para novos achados de pesquisas que possam ser aplicáveis para a nossa própria espécie. Em comparação a outras espécies, os seres humanos são relativamente difíceis de estudar no nível genético e, assim, os avanços na genética humana devem muito a mais de um século de trabalhos com os organismos-modelo.

Todos os organismos-modelo apresentam muito mais do que uma característica útil para os estudos genéticos e outros estudos biológicos. Portanto, após um organismo-modelo ter sido desenvolvido por algumas poucas pessoas com interesses específicos, ele em seguida atua como um núcleo para o desenvolvimento de uma comunidade de pesquisas – um grupo de pesquisadores com um interesse em diversas características de um organismo-modelo em particular. Existem comunidades de pesquisas organizadas para todos os organismos-modelo mencionados neste resumo. Seus integrantes estão regularmente em contato, compartilham suas linhagens mutantes e encontram-se no mínimo uma vez ao ano em conferências que podem atrair milhares de pessoas. Tais comunidades tornam possível a prestação de serviços importantes, como bases de dados de informações de pesquisas, técnicas, estoques genéticos, clones, bibliotecas de DNA e sequências genômicas.

Outra vantagem que pertencer a uma comunidade proporciona a um pesquisador é poder desenvolver uma "percepção do organismo" (expressão da geneticista de milho vencedora de um Prêmio Nobel, Barbara McClintock). Essa ideia é difícil de transmitir, mas implica uma compreensão dos modos gerais de um organismo. Nenhum processo vivo ocorre isolado e, assim, conhecer os modos gerais de um organismo com frequência é benéfico para tentar compreender um processo e interpretá-lo em seu contexto adequado.

À medida que a base de dados em relação a cada organismo-modelo se expande (o que atualmente está ocorrendo a passos largos em virtude da genômica), os geneticistas são cada vez mais capazes de adotar uma visão holística, englobando o funcionamento integrado de todas as partes da constituição do organismo. Desse modo, os organismos-modelo se tornam não apenas modelos para processos isolados, mas também modelos de processos

biológicos integrados. Usa-se a expressão *biologia de sistemas* para descrever essa abordagem holística.

Escherichia coli

Organismo-chave para estudar:

- Transcrição, tradução, replicação, recombinação
- Mutação
- Regulação gênica
- Tecnologia do DNA recombinante.

"Estatística vital" genética

Tamanho do genoma:	4,6 Mb
Cromossomos:	1, circular
Número de genes:	4.000
Porcentagem com homólogos humanos:	8%
Tamanho médio dos genes:	1 kb, nenhum íntron
Transpósons:	Linhagem-específicos, aproximadamente 60 cópias por genoma
Genoma sequenciado em:	1997

A bactéria unicelular *Escherichia coli* é amplamente conhecida como patógeno causador de doenças, fonte de intoxicação alimentar e de doenças intestinais. Entretanto, sua reputação negativa não é merecida. Embora algumas linhagens de *E. coli* sejam prejudiciais, outras são residentes naturais e essenciais do intestino humano. Como organismos-modelo, as linhagens de *E. coli* desempenham um papel indispensável nas análises genéticas. Na década de 1940, diversos grupos começaram a investigar a genética de *E. coli*. Havia necessidade de um organismo simples, que pudesse ser cultivado de modo não dispendioso para produzir grandes quantidades de bactérias individuais, para que fosse possível encontrar e analisar eventos genéticos raros. Tendo em vista que *E. coli* pode ser obtida a partir do intestino humano e é pequena e fácil de cultivar, ela foi uma escolha natural. O trabalho com *E. coli* definiu o início da revelação da "caixa preta" na genética: por meio da seleção e da análise de mutantes, o funcionamento dos processos celulares pôde ser deduzido, ainda que uma célula individual fosse muito pequena para ser visualizada.

Genoma de *E.coli*. Micrografia eletrônica do genoma da bactéria *E. coli*, liberado da célula por meio de

choque osmótico. (G. Murti/Science Source.)

Características especiais

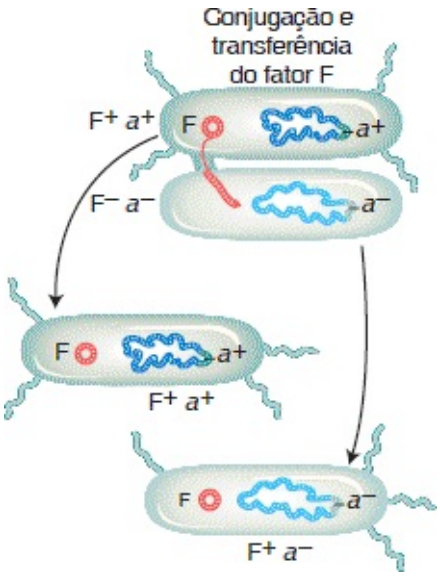
Uma grande parte do sucesso da *E. coli* como organismo-modelo pode ser atribuída a duas estatísticas: seu tamanho celular de 1 µm e um tempo de geração de 20 minutos. (A replicação do cromossomo demora 40 minutos, mas múltiplas forquilhas de replicação possibilitam que a célula se divida em 20 minutos.) Consequentemente, esse procarioto pode crescer em números surpreendentes – uma característica que possibilita que os geneticistas identifiquem mutações e outros eventos genéticos raros, tais como recombinantes intragênicos. A *E. coli* também é notavelmente fácil de cultivar. Quando as células são semeadas em placas de meio com nutriente, cada célula se divide *in situ* e forma uma colônia visível. Alternativamente, grupos de células podem ser cultivados em cultura líquida sob agitação. Fenótipos tais como o tamanho da colônia, a resistência a fármacos, a capacidade de obter energia a partir de fontes de carbono em particular e a produção de corante colorido assumem o lugar dos fenótipos morfológicos da genética de eucariotos.

Colônias bacterianas. (Biophoto Associates/Science Source.)

Ciclo de vida

A *Escherichia coli* se reproduz assexuadamente por meio da fissão celular

simples; seu genoma haploide se replica e se divide com a célula em divisão. Na década de 1940, Joshua Lederberg e Edward Tatum descobriram que *E. coli* também apresenta um tipo de ciclo sexuado, no qual células de "sexos" geneticamente diferenciados se fundem e trocam seus genomas, de forma parcial ou total, por vezes levando à recombinação (ver Capítulo 5). "Machos" conseguem converter "fêmeas" em machos por meio da transmissão de um plasmídio em particular. Esse plasmídio de DNA de 100 kb extragenômico circular, denominado F, determina um tipo de "masculinidade". As células F^+ que atuam como "doadores" machos transmitem uma cópia do plasmídio F para uma célula receptora. O plasmídio F consegue se integrar no cromossomo para formar um tipo de célula Hfr, que transmite o cromossomo linearmente para receptores F^-. Outros plasmídios são observados em *E. coli* na natureza. Alguns carreiam genes cujas funções equipam a célula para a vida em ambientes específicos; os plasmídios R que carreiam genes de resistência a fármacos são exemplos.



Duração do ciclo de vida: 20 minutos.

Geneticistas também têm se beneficiado de alguns elementos genéticos únicos associados à *E. coli*. Plasmídios e fagos bacterianos são utilizados como vetores para clonar os genes de outros organismos em *E. coli*. Os elementos de

transposição de *E. coli* são usados para romper genes em DNA eucariótico clonado. Esses elementos bacterianos são fundamentais na tecnologia do DNA recombinante.

Análise genética

Mutantes espontâneos de *E. coli* demonstram uma diversidade de alterações no DNA, que variam desde simples substituições de bases até a inserção de elementos de transposição. O estudo de mutações espontâneas raras na *E. coli* é possível, uma vez que grandes populações podem ser triadas. Entretanto, mutágenos também são utilizados para aumentar as frequências de mutação.

Para a obtenção de fenótipos mutantes específicos que possam representar defeitos em um processo sob estudo, devem ser planejadas triagens ou seleções. Por exemplo, mutações nutricionais e mutações que conferem resistência a fármacos ou fagos podem ser obtidas em placas suplementadas com substâncias químicas, fármacos ou fagos específicos. Mutações nulas de qualquer gene essencial resultarão em ausência de crescimento; essas mutações podem ser selecionadas por meio da adição de penicilina (um fármaco antibacteriano isolado a partir de um fungo), que mata as células em divisão, mas não os mutantes que não crescem. Em relação às mutações letais condicionais, pode ser utilizado o plaqueamento réplica: colônias que sofreram mutação em uma placa máster são transferidas por meio de uma almofada de feltro para outras placas e, em seguida, são submetidas a algum ambiente tóxico. As mutações que afetam a expressão de um gene de interesse específico podem ser triadas por meio de sua fusão com um gene repórter, tal como o *lacZ*, cujo produto proteico pode produzir um corante azul, ou o gene *GFP*, cujo produto fluoresce quando exposto à luz de um comprimento de onda em particular.

Após um conjunto de mutantes que afetam o processo de interesse ter sido obtido, as mutações são distribuídas em seus genes por meio de recombinação e complementação. Esses genes são clonados e sequenciados para a obtenção de indicações da função. A mutagênese direcionada pode ser utilizada para adequar alterações induzidas por mutação a posições proteicas específicas (ver Capítulo

14).

Em *E. coli*, são utilizados cruzamentos para mapear mutações e produzir genótipos celulares específicos (ver Capítulo 5). Os recombinantes são produzidos por meio da mistura de células Hfr (que apresentam um plasmídio F integrado) e de células F^-. Em geral, um doador Hfr transmite parte do genoma bacteriano, formando um merozigoto temporário, no qual ocorre recombinação. Cruzamentos de Hfr podem ser utilizados para realizar o mapeamento de acordo com o tempo de entrada de marcador ou por meio da frequência de recombinantes. Por meio da transferência de derivados F′ que carreiam genes doadores para F^-, é possível produzir diploides parciais estáveis para estudar a interação gênica ou dominância.

Técnicas de modificação genética

Mutagênese-padrão:

Substâncias químicas e radiação	Mutações somáticas aleatórias
Transpósons	Inserções somáticas aleatórias

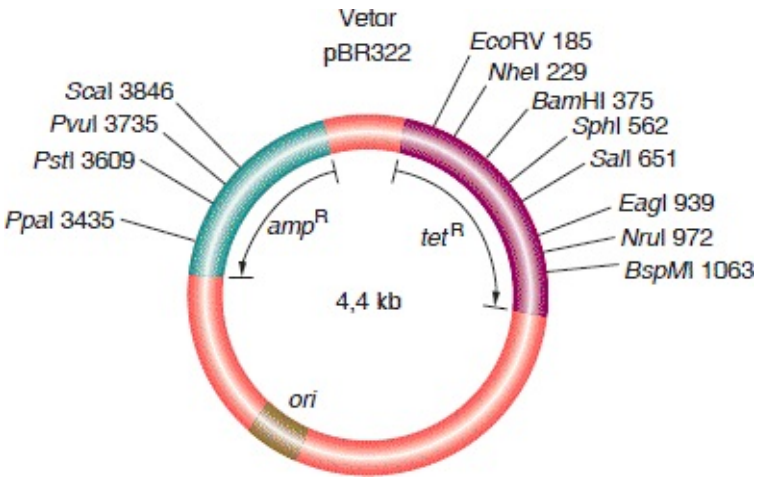
Transgênese:

No plasmídio vetor	Livre ou integrado
No fago vetor	Livre ou integrado
Transformação	Integrada

Nocautes gênicos direcionados:	
Alelo nulo no vetor	Substituição gênica por recombinação
Alelo modificado no vetor	Mutagênese sítio-dirigida por substituição gênica

Engenharia genética

Transgênese. A *E. coli* desempenha um papel-chave na introdução de transgenes em outros organismos (ver Capítulo 10). Ela é o organismopadrão utilizado para a clonagem de genes de qualquer organismo. Os plasmídios ou bacteriófagos de *E. coli* são utilizados como vetores, que carreiam a sequência de DNA a ser clonada. Esses vetores são introduzidos em uma célula bacteriana por meio de transformação, se for um plasmídio, ou por meio de transdução, se for um fago, dentro de cujo citoplasma eles se replicam. Os vetores são especialmente modificados para incluir sítios de clonagem únicos, que podem ser cortados por meio de uma diversidade de enzimas de restrição. Outros vetores "de transporte" são projetados para mover fragmentos de DNA de levedura ("a *E. coli* eucariótica") para *E. coli*, em virtude da sua facilidade de manipulação genética, e, em seguida, de volta à levedura para avaliação fenotípica.



Um plasmídio projetado como um vetor para a clonagem de DNA. A inserção bem-sucedida de um gene exógeno no plasmídio é detectada por meio da inativação de qualquer gene de resistência a fármacos (tet^R ou amp^R). Os sítios de restrição estão identificados.

Nocautes gênicos direcionados. Um conjunto completo de nocautes gênicos está sendo acumulado. Em um procedimento, um transpóson de resistência à canamicina é introduzido em um gene clonado *in vitro* (por meio da utilização de uma transposase). O construto é transformado nele, sendo as colônias resistentes nocautes produzidos por meio de recombinação homóloga.

Contribuições principais

Estudos pioneiros em relação à genética como um todo foram realizados em *E. coli*. Talvez o maior triunfo tenha sido a elucidação do código genético universal de 64 códons, mas essa conquista está muito longe de ser a única na lista de realizações atribuíveis a esse organismo. Outros fundamentos da genética que foram demonstrados pela primeira vez em *E. coli* incluem a natureza espontânea da mutação (o teste de flutuação, Capítulo 16), os diversos tipos de alterações de bases que causam mutações e a replicação semiconservativa do DNA (o experimento de Meselson e Stahl, Capítulo 7). Essa bactéria ajudou a abrir novas áreas da genética, tais como a regulação gênica (o óperon *lac*, Capítulo 11) e a transposição do DNA (elementos IS, Capítulo 15). Por último, mas não menos importante, a tecnologia do DNA recombinante foi inventada em *E. coli* e atualmente o organismo ainda desempenha um papel central nessa tecnologia.

Outras áreas de contribuição

- Metabolismo celular
- Supressores sem sentido
- Colinearidade de gene e polipeptídio
- Óperon
- Resistência a fármacos baseada em plasmídios
- Transporte ativo.

Saccharomyces cerevisiae

Organismo-chave para estudar:

- Genômica
- Biologia de sistemas
- Controle genético do ciclo celular
- Transdução de sinal
- Recombinação
- Tipo reprodutivo
- Herança mitocondrial
- Interação gênica; di-híbrido.

"Estatística vital" genética

Tamanho do genoma:	12 Mb
Cromossomos:	n = 16
Número de genes:	6.000
Porcentagem com homólogos humanos:	25%
Tamanho médio dos genes:	1,5 kb, 0,03 íntron/gene
Transpósons:	Pequena proporção do DNA
Genoma sequenciado em:	1996

O ascomiceto *S. cerevisiae*, conhecido como “levedura do pão”, “broto de levedura” ou simplesmente “levedura”, tem sido a base das indústrias panificadora e cervejeira desde a Antiguidade. Na natureza, provavelmente cresce sobre as superfícies das plantas, utilizando exsudatos como nutrientes, embora o seu nicho preciso ainda seja um mistério. Embora as linhagens de laboratório sejam principalmente haploides, as células na natureza podem ser diploides ou poliploides. Em aproximadamente 70 anos de pesquisas genéticas, a levedura se tornou “*E. coli* dos eucariotos”. Tendo em vista que ela é haploide e unicelular, e que forma colônias compactas em placas, ela pode ser tratada em grande parte do mesmo modo que uma bactéria. Entretanto, ela apresenta meiose, ciclo celular e mitocôndrias eucarióticas, características centrais da história de sucesso da levedura.

Células de levedura, *Saccharomyces cerevisiae*. (SciMAT/Science Source.)

Características especiais

Como um organismo-modelo, a levedura combina o melhor de dois mundos: ela apresenta uma grande parte da conveniência de uma bactéria, mas com as características-chave de um eucarioto. As células de levedura são pequenas (10 µm) e completam seu ciclo celular em apenas 90 minutos, o que possibilita que enormes quantidades sejam produzidas em um curto período. Assim como as bactérias, a levedura pode ser cultivada em grandes grupos em um meio líquido agitado continuamente. E, assim como as bactérias, a levedura produz colônias

visíveis quando semeada em meio com ágar, pode ser triada em relação a mutações e plaqueada em réplica. De maneira eucariótica típica, a levedura apresenta um ciclo de divisão celular mitótica, sofre meiose e contém mitocôndrias que hospedam um pequeno genoma único. As células de levedura podem realizar respiração anaeróbica ao utilizar o ciclo de fermentação e, portanto, sem as mitocôndrias, o que possibilita que os mutantes mitocondriais sejam viáveis.

Análise genética

A realização de cruzamentos em levedura é consideravelmente direta. Linhagens de tipo reprodutivo oposto são simplesmente misturadas em um meio apropriado. Os diploides *a/α* resultantes são induzidos a sofrer meiose mediante a utilização de um meio de esporulação especial. Os investigadores podem isolar os ascósporos de uma tétrade única ao utilizar uma máquina denominada micromanipulador. Eles também têm a opção de sintetizar diploides *a/a* ou α/α para finalidades especiais ou criar diploides parciais por meio da utilização de plasmídios especialmente modificados.

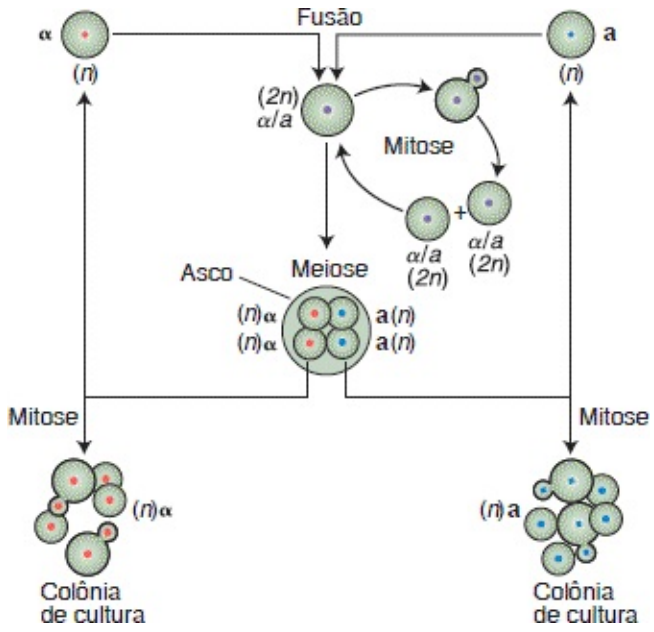
Tendo em vista que está disponível uma enorme variedade de mutantes de leveduras e construtos de DNA na comunidade científica, linhagens para finalidades especiais de triagens e seleções podem ser produzidas por meio do cruzamento de diversos tipos de leveduras. Além disso, novos alelos mutantes podem ser mapeados por meio do cruzamento com linhagens que contêm uma diversidade de marcadores fenotípicos ou de DNA de posição conhecida no mapa.

A disponibilidade de células haploides e diploides proporciona flexibilidade para os estudos mutacionais. As células haploides são convenientes para as seleções ou triagens em grande escala, tendo em vista que os fenótipos mutantes são expressos diretamente. As células diploides são convenientes para obter mutações dominantes, abrigar mutações letais, realizar testes de complementação e explorar interações gênicas.

Ciclo de vida

A levedura é uma espécie unicelular com um ciclo de vida muito simples, composto pelas fases sexuada e assexuada. A fase assexuada pode ser haploide ou diploide. Uma célula se divide assexuadamente por meio de brotamento: uma célula-mãe emite um broto, para o qual é transmitido um dos núcleos que resulta da mitose. Para a reprodução sexuada, existem dois tipos reprodutivos, determinados pelos alelos *MATα* e *MATa*. Quando células haploides de diferentes tipos reprodutivos se unem, elas formam uma célula diploide, que pode se dividir mitoticamente ou sofrer divisão meiótica. O produto da meiose é uma tétrade não linear de quatro ascósporos.

Duração total do ciclo de vida: 90 minutos para completar o ciclo celular.



Técnicas de manipulação genética

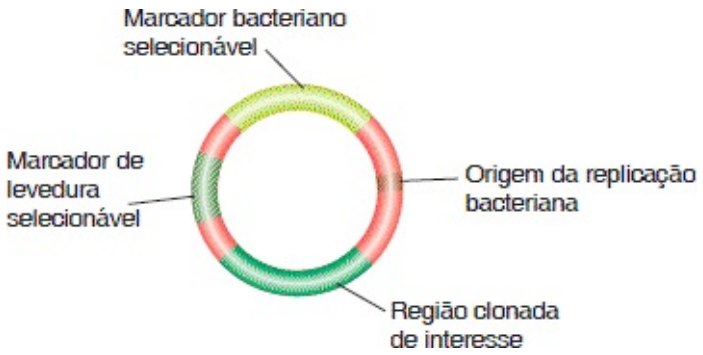
Mutagênese-padrão:

Substâncias químicas e radiação	Mutações somáticas aleatórias
Transpósons	Inserções somáticas aleatórias
Transgênese:	
Plasmídio integrativo	Inserções por recombinação homóloga
Plasmídio replicativo	Pode se replicar de modo autônomo (origem de replicação 2μ ou ARS)
Cromossomo artificial de levedura	Replica e segrega como um cromossomo
Vetor de transporte	Pode replicar em levedura ou *E. coli*
Nocautes gênicos direcionados:	
Substituição de gene	A recombinação homóloga substitui o alelo do tipo selvagem por uma cópia nula

Engenharia genética

Transgênese. A levedura em brotamento proporciona mais oportunidades para a manipulação genética do que qualquer outro eucarioto (ver Capítulo 10). O DNA

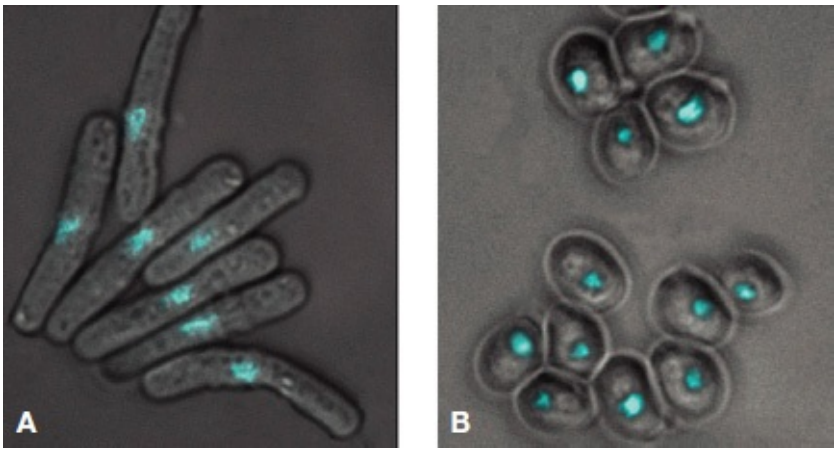
exógeno é facilmente captado pelas células cujas paredes celulares foram parcialmente removidas por meio de digestão enzimática ou abrasão. Estão disponíveis diversos tipos de vetores. Para que um plasmídio se replique livre dos cromossomos, ele deve conter uma origem normal de replicação em levedura (ARS) ou uma origem de replicação de um plasmídio de 2 µm encontrado em determinados isolados de levedura. O vetor mais elaborado, o cromossomo artificial de levedura (YAC), é composto por um ARS, um centrômero de levedura e dois telômeros. Um YAC pode carrear grandes insertos transgênicos, que em seguida são herdados do mesmo modo que os cromossomos mendelianos. Os YAC têm sido vetores importantes na clonagem e no sequenciamento de genomas grandes, tais como o genoma humano.



Um vetor de levedura simples. Este tipo de vetor é denominado plasmídio integrativo de levedura (YIp).

Nocautes direcionados. A mutagênese de transpóson (marcação de transpóson) pode ser alcançada por meio da introdução do DNA de levedura em *E. coli* em um vetor de transporte; os transpósons bacterianos são integrados ao DNA da levedura, provocando nocaute (inativação) da função gênica. O vetor de transporte em seguida é transferido de volta para a levedura e os mutantes marcados substituem as cópias do tipo selvagem por meio de recombinação homóloga. Os nocautes gênicos também podem ser conquistados por meio da substituição dos alelos do tipo selvagem por uma cópia nula modificada, por meio de recombinação homóloga. Ao utilizar essas técnicas, pesquisadores construíram sistematicamente um conjunto completo de linhagens nocauteadas de levedura (cada uma carreando um nocaute diferente) para avaliar a função nula de

cada gene no nível fenotípico.



Mutantes de ciclo celular. A. Mutantes que se alongam sem se dividir. **B.** Mutantes que param sem brotamento. (Cortesia de Susan L. Forsburg, the Salk Institute. "The Art and Design of Genetic Screens: Yeast", *Nature Reviews: Genetics 2*, 2001, 659-668.)

Contribuições principais

Graças a uma combinação de genética adequada e bioquímica adequada, estudos em leveduras realizaram contribuições substanciais para a nossa compreensão sobre o controle genético dos processos celulares.

Ciclo celular. A identificação de genes de divisão celular por meio de seus mutantes sensíveis à temperatura (mutantes *cdc*) levou a um modelo poderoso para o controle genético da divisão celular. Os diferentes fenótipos Cdc revelam os componentes do maquinário necessário para a execução de etapas específicas na progressão do ciclo celular. Esse trabalho tem sido útil para a compreensão sobre os controles da divisão celular anormal que podem levar ao câncer humano.

Recombinação. Muitas das ideias-chave em relação aos atuais modelos moleculares de *crossing over* (tais como o modelo de quebra bifilamentar) têm por base a análise de tétrades da conversão gênica em leveduras (ver Capítulo 4). A conversão gênica (proporções alélicas aberrantes, tais como 3:1) é consideravelmente comum nos genes de levedura, proporcionando um conjunto de dados adequadamente grande para a quantificação das características-chave desse

processo.

Interações gênicas. A levedura liderou o caminho no estudo das interações gênicas. As técnicas da genética tradicional têm sido utilizadas para revelar padrões de epistasia e supressão, que sugerem interações gênicas (ver Capítulo 6). O sistema de plasmídio di-híbrido para a observação das interações proteicas foi desenvolvido em levedura e gerou mapas de interação complexos, que representam o início da biologia de sistemas (ver Capítulo 14). Letais sintéticos–mutantes duplos letais criados por meio do intercruzamento de dois mutantes únicos viáveis também são utilizados para criar redes de interação (ver Capítulo 6).

Genética mitocondrial. Mutantes com mitocôndrias defeituosas são reconhecíveis como colônias muito pequenas, denominadas "petites". A disponibilidade dessas petites e de outros mutantes mitocondriais possibilitou a primeira análise detalhada da estrutura e da função do genoma mitocondrial em qualquer organismo.

Genética do tipo reprodutivo. Os alelos *MAT* de leveduras foram os primeiros genes de tipos reprodutivos a ser caracterizados no nível molecular. Curiosamente, a levedura é submetida à alteração espontânea de um tipo reprodutivo para outro. Uma cópia "sobressalente" silenciosa do alelo *MAT* oposto, localizada em algum outro local no genoma, entra no *locus* do tipo reprodutivo, substituindo o alelo residente por meio de recombinação homóloga. A levedura proporcionou um dos modelos centrais para a transdução de sinais durante a detecção e a resposta aos hormônios reprodutivos do tipo reprodutivo oposto.

Outras áreas de contribuição

- Genética da alteração entre o crescimento semelhante à levedura e o filamentoso
- Genética da senescência.

Neurospora crassa

Organismo-chave para estudar:

- Genética do metabolismo e da captação
- Genética do *crossing over* e da meiose
- Citogenética de fungos
- Crescimento polar
- Ritmos circadianos
- Interações do núcleo com as mitocôndrias.

"Estatística vital" genética

Tamanho do genoma:	43 Mb
Cromossomos:	7 autossomos (n = 7)
Número de genes:	10.000
Porcentagem com homólogos humanos:	6%
Tamanho médio dos genes:	1,7 kb, 1,7 íntron/gene
Transpósons:	Raros
Genoma sequenciado em:	2003

O *Neurospora crassa*, fungo alaranjado do pão, foi um dos primeiros

microrganismos eucarióticos a ser adotado pelos geneticistas como um organismo-modelo. Assim como a levedura, originalmente ele foi escolhido em virtude de sua haploidia, de seu ciclo de vida simples e rápido, e da facilidade com a qual ele pode ser cultivado. Particularmente significativo foi o fato de que ele cresce em um meio com um conjunto definido de nutrientes, tornando possível estudar o controle genético da química celular. Na natureza, ele é observado em muitas partes do mundo, crescendo sobre a vegetação morta. Em virtude da ativação dos seus ascósporos quiescentes pelo fogo, ele é mais facilmente coletado após queimadas – por exemplo, sob a casca de árvores queimadas e em campos de cultivos tais como os de cana-de-açúcar, que são queimados de modo rotineiro antes da coleta.

***Neurospora crassa* crescendo em cana-de-açúcar.** (Cortesia de David Jacobson.)

Características especiais

Neurospora detém o recorde de velocidade dos fungos, tendo em vista que cada hifa cresce mais de 10 cm ao dia. Esse rápido crescimento, combinado com o seu ciclo de vida haploide e a capacidade de crescer em meio definido, tornou-o um organismo de escolha para o estudo da genética bioquímica da nutrição e da captação de nutrientes.

Outra característica única de *Neurospora* (e de fungos relacionados) possibilita que geneticistas tracem as etapas de meioses únicas. Os quatro produtos haploides de uma meiose permanecem unidos em um saco, denominado asco. Cada um dos quatro produtos da meiose sofre a uma divisão mitótica adicional, resultando em uma óctade linear de oito ascósporos (ver Capítulo 3).

Essa característica torna o *Neurospora* um sistema ideal para o estudo de *crossing over*, conversão gênica, rearranjos cromossômicos, ausência de disjunção meiótica e controle genético da própria meiose. Os cromossomos, embora pequenos, são facilmente visíveis e, assim, os processos meióticos podem ser estudados nos níveis genético e cromossômico. Portanto, em *Neurospora*, foram realizados estudos fundamentais sobre os mecanismos subjacentes a esses processos (ver Capítulo 4).

Análise genética

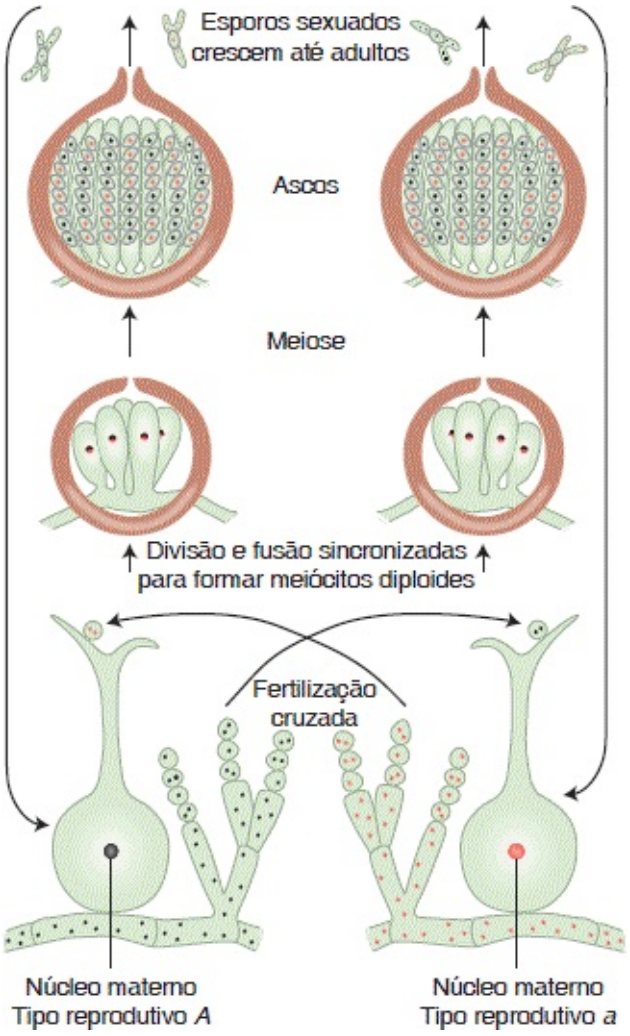
A análise genética é direta (ver Capítulo 3). Centros que mantêm estoques fornecem uma ampla diversidade de mutantes que afetam todos os aspectos da biologia do fungo. Os genes de *Neurospora* podem ser facilmente mapeados por meio de seu cruzamento com um banco de linhagens com *loci* mutantes ou alelos de RFLP conhecidos. Linhagens de tipo reprodutivo oposto são cruzadas simplesmente por meio do seu cultivo em conjunto. Um geneticista com uma agulha pode coletar um único ascósporo para estudo. Portanto, as análises nas quais são utilizados ascos completos ou ascósporos aleatórios são rápidas e diretas.

Ciclo de vida

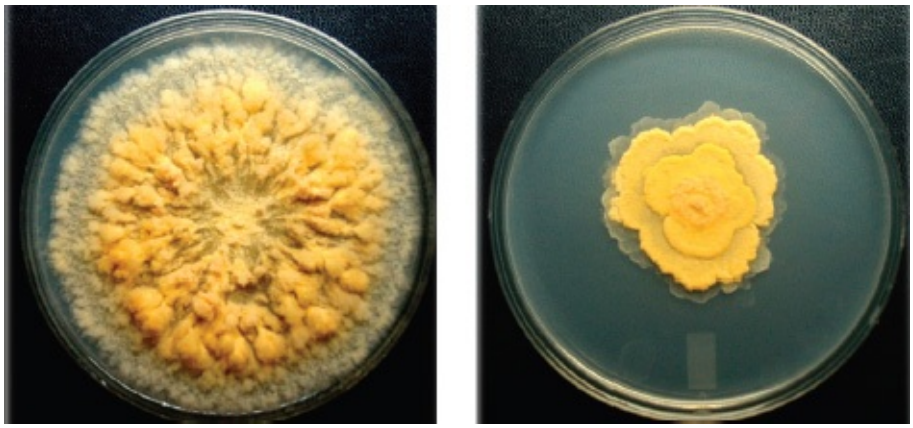
N. crassa apresenta um ciclo de vida eucariótico haploide. Um esporo assexuado haploide (denominado conídio) germina para produzir um tubo germinativo que se estende na sua ponta. O crescimento progressivo da ponta e a ramificação produzem uma massa de filamentos ramificados (denominados *hifas*), que forma uma colônia compacta no meio de crescimento. Tendo em vista que as hifas não apresentam paredes transversais, uma colônia é essencialmente uma célula que contém muitos núcleos haploides. A colônia brota e libera milhões de esporos assexuados, que podem se dispersar no ar e repetir o ciclo assexuado.

No ciclo sexuado de *N. crassa*, existem dois tipos reprodutivos de aspecto

idêntico, MAT-*A* e MAT-*a*, que podem ser considerados “sexos” simples. Assim como em leveduras, os dois tipos reprodutivos são determinados pelos dois alelos de um gene. Quando colônias de tipos reprodutivos diferentes entram em contato, as suas paredes celulares e os seus núcleos se fundem. Surgem muitos núcleos diploides temporários, cada um sofrendo meiose, produzindo uma óctade de ascósporos. Os ascósporos germinam e produzem colônias exatamente como aquelas produzidas pelos esporos assexuados.



Duração do ciclo de vida: 4 semanas para o ciclo sexuado.



***Neurospora* do tipo selvagem (*esquerda*) e mutante (*direita*) cultivados em uma placa de Petri.** (Cortesia de Anthony Griffiths/Olivera Gavric.)

Tendo em vista que *Neurospora* é haploide, fenótipos mutantes recém-obtidos são facilmente detectados com a utilização de diversos tipos de triagem e seleção. Um sistema favorito para o estudo do mecanismo de mutação é o gene *ad-3*, tendo em vista que mutantes do *ad-3* são roxos e facilmente detectados.

Embora diploides vegetativos de *Neurospora* não sejam prontamente obtidos, geneticistas são capazes de criar um "diploide mimético", útil para testes de complementação e outras análises que requerem a presença de duas cópias de um gene (ver Capítulo 6). A saber, a fusão de duas linhagens diferentes produz um heterocário, indivíduo que contém dois tipos nucleares diferentes em um citoplasma comum. Os heterocários também possibilitam a utilização de uma versão do teste de *locus* específico, um modo de recuperar mutações em um alelo recessivo específico. (As células de um heterocário +/*m* são plaqueadas e busca-se pelas colônias *m*/*m*.)

Técnicas de manipulação genética

Mutagênese-padrão:

Substâncias químicas e radiação	Mutações somáticas aleatórias

Mutagênese de transpósons	Não disponível
Transgênese:	
Transformação mediada por plasmídio	Inserção aleatória
Nocautes de gênicos direcionados:	
RIP	Mutações de GC → AT em segmentos transgênicos duplicados antes de um cruzamento
Repressão	Inativação pós-transcricional somática de transgenes

Engenharia genética

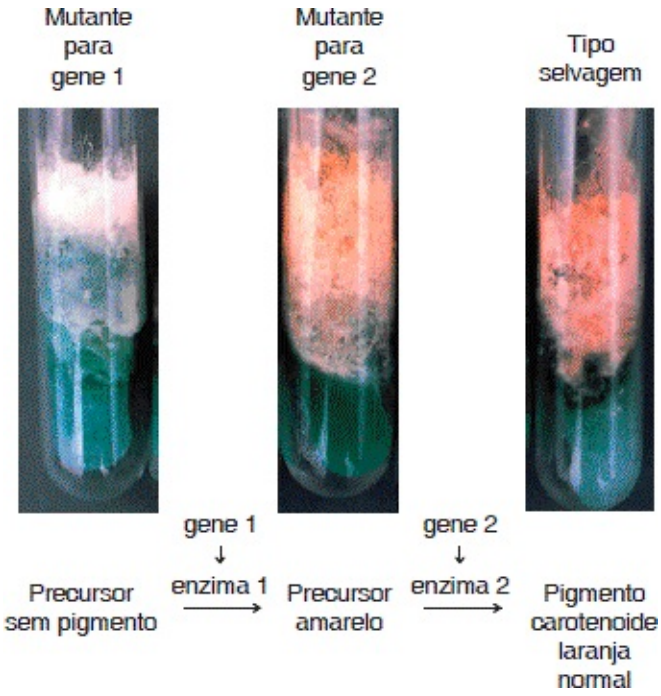
Transgênese. A primeira transformação eucariótica foi conquistada em *Neurospora*. Atualmente, *Neurospora* é facilmente transformado com a utilização de plasmídios bacterianos que carreiam o transgene desejado, mais um marcador selecionável, tal como a resistência à higromicina, para demonstrar que o plasmídio entrou. Nenhum plasmídio replica em *Neurospora* e, assim, um transgene é herdado apenas se estiver integrado a um cromossomo.

Nocautes direcionados. Em linhagens especiais de *Neurospora*, os transgenes com frequência integram-se por meio de recombinação homóloga. Portanto, uma linhagem transgênica normalmente apresenta o gene residente, mais o transgene

homólogo, inserido em uma localização ectópica aleatória. Em virtude dessa duplicação de material, se a linhagem for cruzada, ela está sujeita a RIP, um processo genético que é único de *Neurospora*. RIP é um mecanismo pré-meiótico que introduz muitas transições de GC para AT em ambas as cópias duplicadas, rompendo efetivamente o gene. Portanto, RIP pode ser aproveitado como um modo conveniente para nocautear deliberadamente um gene específico.

Contribuições principais

George Beadle e Edward Tatum utilizaram *Neurospora* como organismo-modelo em seus estudos pioneiros sobre as relações entre genes e enzimas, nos quais foram capazes de determinar as etapas enzimáticas na síntese da arginina (ver Capítulo 6). Seu trabalho com *Neurospora* estabeleceu o início da genética molecular. Seguiram-se muitos estudos comparáveis sobre a genética do metabolismo celular com a utilização de *Neurospora*.



Via de síntese do pigmento carotenoide laranja em *Neurospora*. (Cortesia de Anthony Griffiths.)

Têm sido realizados trabalhos pioneiros sobre a genética dos processos meióticos, tais como *crossing over* e disjunção, e sobre os ritmos de formação de

conídios. Culturas em crescimento contínuo demonstram um ritmo diário de formação de conidiósporos. Os resultados de estudos pioneiros com a utilização de mutações que alteram esse ritmo contribuíram para um modelo geral em relação à genética dos ritmos circadianos.

O *Neurospora* atua como um modelo em relação à grande quantidade de fungos filamentosos patogênicos que afetam cultivos e seres humanos, tendo em vista que esses fungos com frequência são difíceis de cultivar e manipular geneticamente. Ele é até mesmo utilizado como um sistema de teste eucariótico simples para substâncias mutagênicas e carcinogênicas no ambiente humano.

Tendo em vista que podem ser realizados cruzamentos com a utilização de apenas uma genitora, o ciclo é conveniente para o estudo da genética mitocondrial e da interação de núcleo e mitocôndrias. Foi descoberta uma ampla variedade de plasmídios mitocondriais lineares e circulares em isolados naturais. Alguns deles são retroelementos que se acredita serem intermediários na evolução dos vírus.

Outras áreas de contribuição

- Diversidade e adaptação dos fungos
- Citogenética (base cromossômica da genética)
- Genes do tipo reprodutivo
- Genes de compatibilidade de heterocário (um modelo para o reconhecimento genético próprio e não próprio).

Arabidopsis thaliana

Organismo-chave para estudar:

- Desenvolvimento
- Expressão e regulação gênica
- Genômica de plantas.

"Estatística vital" genética

Tamanho do genoma:	125 Mb
Cromossomos:	Diploide, 5 autossomos ($2n = 10$)
Número de genes:	25.000
Porcentagem com homólogos humanos:	18%
Tamanho médio dos genes:	2 kb, 4 íntrons/gene
Transpósons:	10% do genoma
Genoma sequenciado em:	2000

A *Arabidopsis thaliana*, um membro da família Brassicaceae (repolho, couve) de plantas, chegou de modo relativamente tardio como um organismo-modelo genético. A maior parte dos trabalhos foi realizada nos últimos 20 anos. Ela não apresenta significância econômica: ela cresce de modo prolífico como uma erva daninha em muitas partes temperadas do mundo. Entretanto, em virtude do seu pequeno tamanho, do ciclo de vida curto, e do genoma pequeno, ela ultrapassou os modelos genéticos de plantas mais tradicionais, tais como o milho e o trigo, e se tornou o modelo dominante para a genética molecular vegetal.

Arabidopsis thaliana crescendo na natureza. As versões cultivadas em laboratório são menores. (Floral Images/Alamy.)

Características especiais

Em comparação a outras plantas, a *Arabidopsis* é pequena em relação ao seu tamanho físico e ao tamanho de seu genoma – características que são vantajosas para um organismo-modelo. A *Arabidopsis* cresce até uma altura inferior a 10 cm sob condições apropriadas; portanto, pode ser cultivada em grandes quantidades, possibilitando triagens de mutantes e análises de progênie em grande escala. O tamanho total de seu genoma de 125 Mb tornou o genoma relativamente fácil de sequenciar em comparação a outros genomas de organismos-modelo de plantas, tais como o genoma do milho (2.500 Mb) e o genoma do trigo (16.000 Mb).

Análise genética

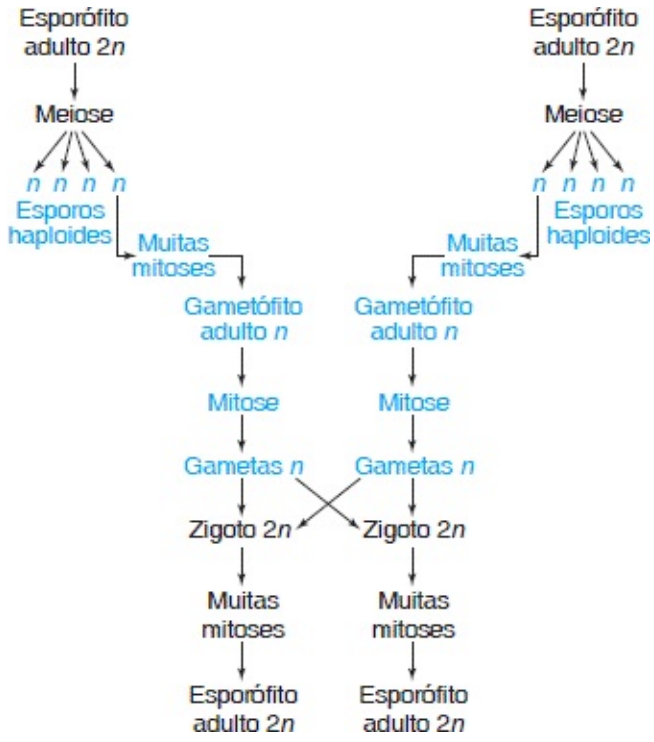
A análise de mutações em *Arabidopsis* por meio de cruzamentos depende de tentativas e métodos verdadeiros – essencialmente aqueles utilizados por Mendel. Os estoques de plantas que carreiam mutações úteis e relevantes para o experimento são obtidos a partir de centros públicos. As linhagens podem ser

cruzadas manualmente umas com as outras ou podem ser autofertilizadas. Embora as flores sejam pequenas, a polinização cruzada é facilmente realizada por meio da remoção das anteras que não se abriram (que por vezes são ingeridas pelo experimentador como um meio conveniente de descarte). Cada flor polinizada em seguida produz uma vagem longa, que contém um grande número de sementes. Essa abundante produção de descendência (milhares de sementes por planta) é uma vantagem para os geneticistas que procuram por mutantes raros ou outros eventos raros. Se uma planta carreia uma nova mutação recessiva na linhagem germinativa, a autofecundação possibilita que a progênie homozigota para a mutação recessiva seja recuperada nos descendentes imediatos da planta.

Ciclo de vida

A *Arabidopsis* apresenta o ciclo de vida familiar das plantas, com um estágio diploide dominante. Uma planta contém diversas flores, cada qual produzindo muitas sementes. Assim como muitas ervas daninhas anuais, o seu ciclo de vida é rápido: demora apenas aproximadamente 6 semanas para que uma semente plantada produza uma nova colheita de sementes.

Duração total do ciclo de vida: 6 semanas.



Mutantes de *Arabidopsis*. (*Esquerda*) Flor de *Arabidopsis* do tipo selvagem. (*Centro*) Mutação *agamous* (*ag*), que resulta em flores apenas com pétalas e sépalas (nenhuma estrutura reprodutiva). (*Direita*) Um mutante duplo *ap-1*, *cal*, que produz uma flor que se assemelha a uma couve-flor. (Mutações semelhantes no repolho provavelmente são a causa das couves-flores reais.) (George Haughn.)

Técnicas de modificação genética

Mutagênese-padrão:

Substâncias químicas e radiação	Mutações nas linhagens germinativas ou somáticas aleatórias
O próprio T-DNA ou transpósons	Inserções aleatórias marcadas
Transgênese:	
O T-DNA carreia o transgene	Inserção aleatória
Nocautes gênicos direcionados:	
Mutagênese mediada por T-DNA ou transpóson	Inserção aleatória; nocautes de mutagênese selecionados com PCR
RNAi	Simula o nocaute direcionado

Engenharia genética

Transgênese. A T-DNA de *Agrobacterium* é um vetor conveniente para a introdução de transgenes (ver Capítulo 10). O construto vetor-transgene se insere aleatoriamente no genoma. A transgênese oferece um modo efetivo para estudar a regulação gênica. O transgene é unido a um gene repórter, tal como GUS, que produz um corante azul em quaisquer posições na planta na qual o gene estiver ativo.

Nocautes direcionados. Tendo em vista que a recombinação homóloga é rara em *Arabidopsis*, genes específicos não podem ser facilmente nocauteados por meio

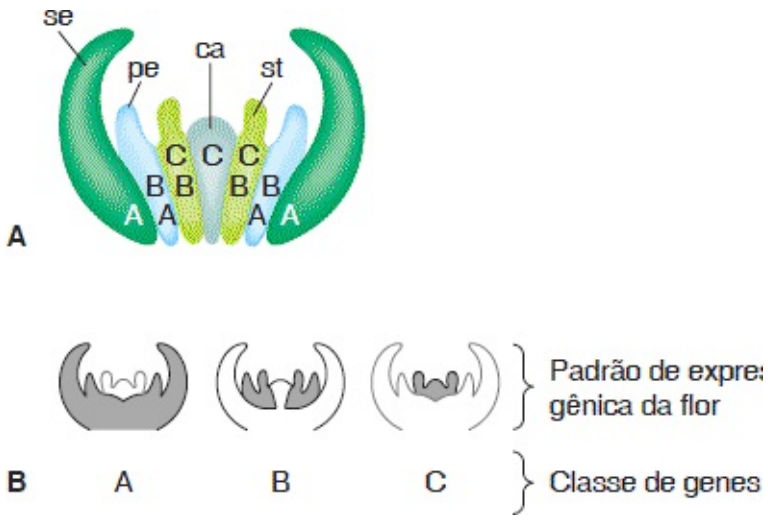
da substituição homóloga com um transgene. Portanto, em *Arabidopsis*, os genes são nocauteados por meio da inserção aleatória de um T-DNA vetor ou transpóson (são utilizados transpósons de milho, tais como *Ac-Ds*), e em seguida nocautes gênicos específicos são selecionados por meio da aplicação da análise por PCR ao DNA de grandes conjuntos de plantas. A PCR utiliza uma sequência no T-DNA ou no transpóson como um *primer* e uma sequência no gene de interesse como o outro *primer*. Portanto, a PCR amplifica apenas as cópias do gene de interesse que carreiam uma inserção. A subdivisão do conjunto e a repetição do processo levam à planta específica que carreia o nocaute. Alternativamente, RNAi pode ser utilizado para inativar um gene específico.

Estão disponíveis grandes coleções de mutantes de inserção de T-DNA; elas apresentam as sequências flanqueadoras das plantas listadas em bases de dados públicas; assim, se você estiver interessado em um gene específico, poderá verificar se a coleção contém uma planta que apresenta uma inserção naquele gene. Uma característica conveniente de populações nocaute em plantas é que elas podem ser mantidas facilmente e de modo não dispendioso como coleções de sementes durante muitos anos, talvez mesmo décadas. Essa característica não é possível em relação à maior parte das populações de modelos animais. O verme *Caenorhabditis elegans* pode ser preservado congelado, mas a mosca-das-frutas (*Drosophila melanogaster*) não pode ser congelada e reavivada. Portanto, linhagens de mutantes de mosca-das-frutas devem ser mantidas vivas.

Contribuições principais

Como o primeiro genoma de planta a ser sequenciado, *Arabidopsis* proporcionou um importante modelo em relação à arquitetura e à evolução do genoma de plantas. Além disso, estudos conduzidos em *Arabidopsis* deram contribuições-chave para a nossa compreensão sobre o controle genético do desenvolvimento das plantas. Geneticistas isolaram mutações homeóticas que afetam o desenvolvimento das flores, por exemplo. Nesses mutantes, um tipo de parte floral é substituído por outro. A integração da ação desses mutantes levou a um modelo sofisticado de determinação do verticilo das flores com base nos padrões

sobrepostos de expressão de genes reguladores no meristema da flor. *Arabidopsis* também contribuiu amplamente para a base genética da fisiologia vegetal, da regulação gênica e da interação das plantas com o ambiente (incluindo a genética da resistência a doenças). Tendo em vista que *Arabidopsis* é uma planta natural de distribuição mundial, ela apresenta um grande potencial para o estudo da diversificação e da adaptação evolutiva.



O estabelecimento do destino do verticilo. A. Padrões de expressão gênica correspondentes aos diferentes destinos do verticilo. Do mais externo para o mais interno, os destinos são sépala (se), pétala (pe), estame (st) e carpelo (ca). **B.** As regiões sombreadas dos diagramas transversais da flor em desenvolvimento indicam os padrões de expressão gênica em relação aos genes das classes A, B e C.

Outras áreas de contribuição

- Resposta ao estresse ambiental
- Sistemas de controle hormonal.

Caenorhabditis elegans

Organismo-chave para estudar:

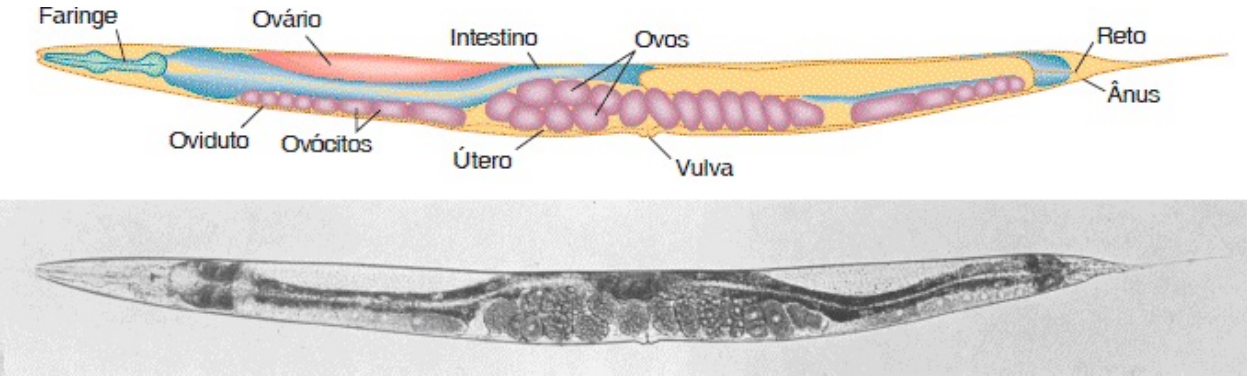
- Desenvolvimento

- Comportamento
- Nervos e músculos
- Envelhecimento.

"Estatística vital" genética

Tamanho do genoma:	97 Mb
Cromossomos:	5 autossomos ($2n$ = 10), cromossomo X
Número de genes:	19.000
Porcentagem com homólogos humanos:	25%
Tamanho médio dos genes:	5 kb, 5 éxons/gene
Transpósons:	Diversos tipos, ativos em algumas linhagens
Genoma sequenciado em:	1998

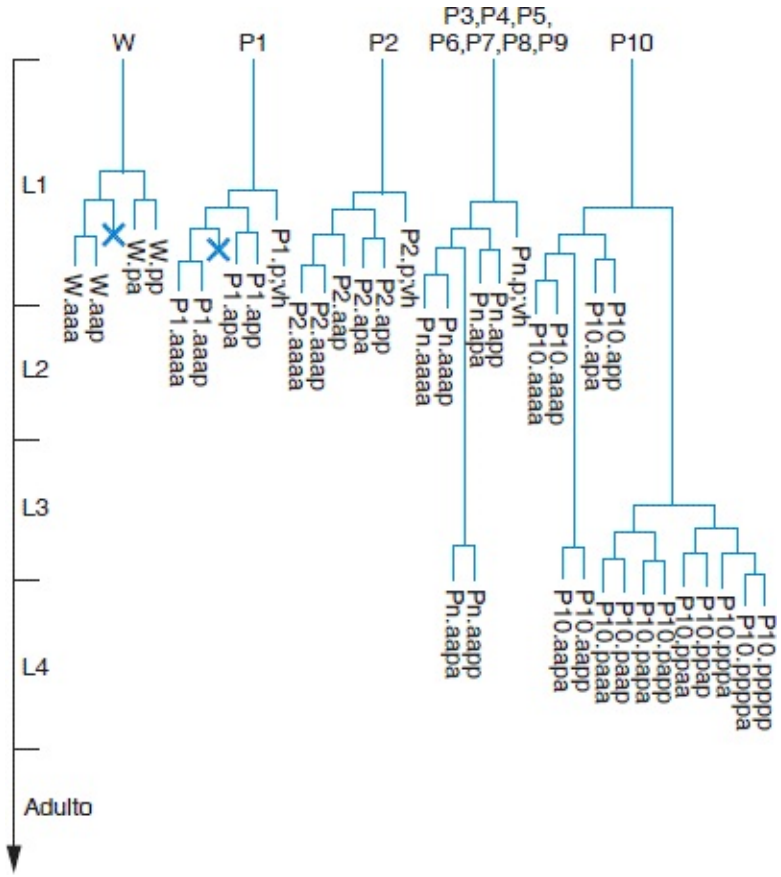
O *Caenorhabditis elegans* pode não parecer grande coisa sob um microscópio e, de fato, esse nematódeo (geo-helminto), de 1 mm de comprimento e que habita o solo, é um animal relativamente simples. Mas tal simplicidade é parte do que torna o *C. elegans* um bom organismo-modelo. Seu tamanho pequeno, o crescimento rápido, a capacidade de autofecundação, a transparência e a baixa quantidade de células no corpo tornaram-no uma escolha ideal para o estudo da genética do desenvolvimento eucariótico.



Fotomicrografia e desenho de um *Caenorhabditis elegans* adulto. (De J. E. Sulston e H. R. Horvitz, *Developmental Biology*, 56, 1977, 111.)

Características especiais

Os geneticistas conseguem enxergar através do *C. elegans*. Contrariamente a outros organismos-modelo multicelulares, tais como a mosca-das-frutas e *Arabidopsis*, esse pequeno nematódeo é transparente, o que o torna eficiente para triar grandes populações em relação a mutações interessantes que afetam praticamente qualquer aspecto da anatomia ou do comportamento. A transparência também o torna apropriado para estudos do desenvolvimento: os pesquisadores conseguem observar diretamente todos os estágios do desenvolvimento ao simplesmente olhar os nematódeos sob um microscópio óptico. Os resultados de tais estudos demonstraram que o desenvolvimento do *C. elegans* é firmemente programado e que cada nematódeo apresenta um número de células surpreendentemente pequeno e consistente (959 nos hermafroditas e 1.031 nos machos). De fato, biólogos rastrearam os destinos de células específicas na medida em que o nematódeo se desenvolve e determinaram o padrão exato de divisão celular que leva a cada órgão adulto. Esse esforço produziu um heredograma de linhagem em relação a todas as células adultas (ver Capítulo 13).



Uma representação simbólica das linhagens de 11 células. Uma célula que é submetida à morte celular programada é indicada por um X azul no final da ramificação de uma linhagem.

Ciclo de vida

C. elegans é único entre os principais animais-modelo, pois um dos dois sexos é hermafrodita (XX). O outro é masculino (XO). Os dois sexos podem ser distinguidos pelo tamanho maior dos hermafroditas e por diferenças nos seus órgãos sexuais. Os hermafroditas produzem ovócitos e espermatozoides e, assim, podem se autofecundar. A progênie de um hermafrodita autofecundado também é hermafrodita, exceto quando uma rara não disjunção origina um macho XO. Se hermafroditas e machos forem misturados, os sexos copulam e muitos dos zigotos resultantes terão sido fertilizados pelo espermatozoide ameboide dos machos. A fertilização e a produção do embrião ocorrem dentro do hermafrodita, que, em seguida, deposita os ovos. Os ovos finalizam o seu

desenvolvimento externamente.

Duração total do ciclo de vida: 3 dias e $\frac{1}{2}$

Análise genética

Tendo em vista que esses nematódeos são pequenos e se reproduzem rapidamente e de modo prolífico (a autofecundação produz aproximadamente uma progênie de 300 organismos e o cruzamento produz aproximadamente 1.000), originam progênie grande, que pode ser triada em relação a eventos genéticos raros. Além disso, tendo em vista que o hermafroditismo no *C. elegans* torna a autofecundação possível, os nematódeos com mutações homozigotas recessivas podem ser rapidamente isolados por meio de autofecundação da progênie tratada. Contrariamente, outros modelos animais, tais como moscas-das-frutas ou camundongos, requerem o cruzamento entre irmãos e demoram mais gerações para isolar mutações recessivas.

Técnicas de modificação genética

Mutagênese-padrão:

Substâncias químicas (EMS) e radiação	Mutações aleatórias na linhagem germinativa
Transpósons	Inserções aleatórias na linhagem germinativa

Transgênese:

Injeção de transgene na gônada	Transgene não integrado; integração

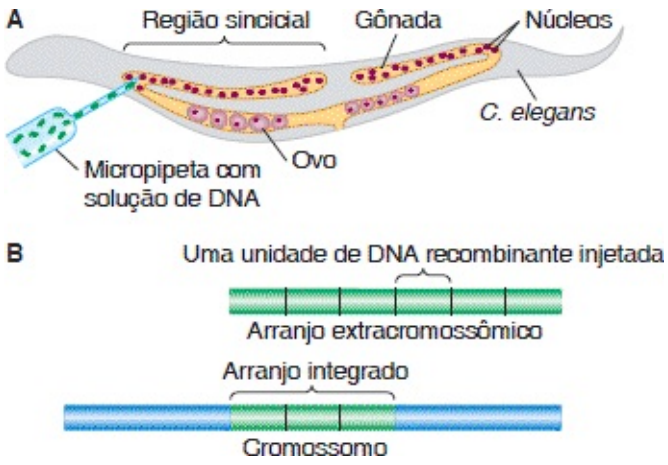
	ocasional
Nocautes gênicos direcionados:	
Mutagênese mediada por transpóson	Nocautes selecionados com PCR
RNAi	Simula o nocaute direcionado
Ablação a *laser*	Nocaute de uma célula

Engenharia genética

Transgênese. A introdução de transgenes na linhagem germinativa é possibilitada por meio de uma propriedade especial das gônadas do *C. elegans*, que são sinciciais, o que significa que existem muitos núcleos em um citoplasma comum. Os núcleos não se tornam incorporados nas células até a meiose, quando a formação de ovócitos individuais ou espermatozoides tem início. Portanto, uma solução de DNA que contém o transgene injetado na gônada de um hermafrodita expõe mais de 100 núcleos precursores de células germinativas ao transgene. Ao acaso, alguns desses núcleos incorporarão o DNA (ver Capítulo 10).

Os transgenes se recombinam para formar arranjos em *tandem* de multicópias. Em um ovócito, os arranjos não se integram em um cromossomo, mas os transgenes dos arranjos ainda são expressos. Portanto, o gene carreado em um clone de DNA do tipo selvagem pode ser identificado por meio da sua introdução em uma linhagem receptora recessiva específica (complementação funcional). Em alguns casos, porém não em todos, os arranjos transgênicos são transmitidos para a progênie. Para aumentar a chance de herança, os vermes são expostos à radiação ionizante, que pode induzir a integração de um arranjo em uma posição cromossômica ectópica e, nesse sítio, o arranjo é transmitido de modo confiável para a progênie.

Nocautes direcionados. Em linhagens com transpósons ativos, os próprios transpósons se tornam agentes de mutação por meio da inserção em locais aleatórios no genoma, nocauteando os genes interrompidos. Se pudermos identificar organismos com inserções em um gene de interesse específico, podemos isolar um nocaute gênico direcionado. Insertos em genes específicos podem ser detectados por meio da utilização da PCR se um *primer* de PCR tiver por base a sequência do transpóson e outro tiver por base a sequência do gene de interesse. Alternativamente, RNAi pode ser utilizado para anular a função de genes específicos. Como uma alternativa para a mutação, células individuais podem ser mortas por um feixe de *laser* para a observação do efeito sobre a função ou o desenvolvimento do verme (ablação a *laser*).



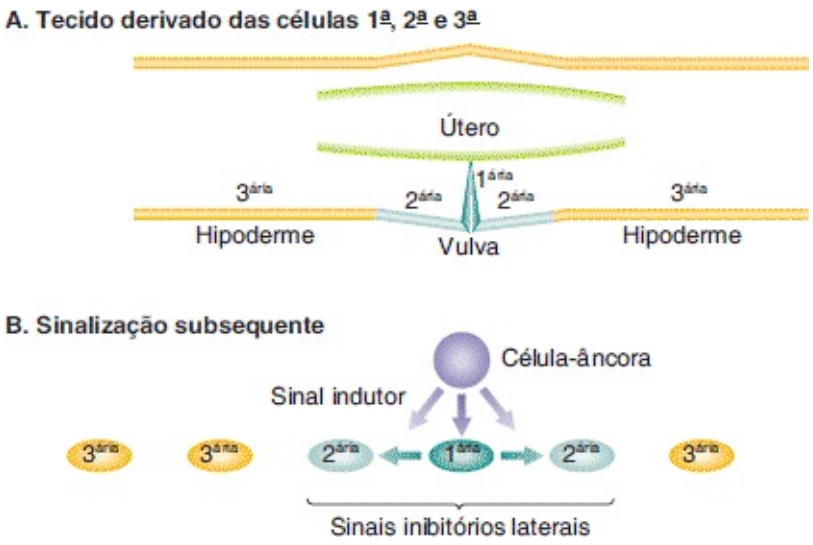
Criação de transgenes de *C. elegans*. A. Método de injeção. **B.** Arranjos extracromossômicos e integrados.

Contribuições principais

O *C. elegans* se tornou um organismo-modelo favorito para o estudo de diversos aspectos do desenvolvimento, em virtude de seu tamanho pequeno e número invariável de células. Um exemplo é a morte celular programada, um aspecto crucial do desenvolvimento normal. Algumas células são geneticamente programadas para morrer no período do desenvolvimento (um processo denominado apoptose). Os resultados dos estudos de *C. elegans* contribuíram para um modelo geral útil em relação à apoptose, que sabidamente também é uma

característica do desenvolvimento humano.

Outro sistema-modelo é o desenvolvimento da vulva, a abertura do trato reprodutivo para o exterior. Hermafroditas com vulvas defeituosas ainda produzem progênie, que em triagens são facilmente visíveis agrupadas dentro do corpo. Os resultados de estudos de hermafroditas sem vulvas ou com muitas revelaram como as células que apresentam um início completamente equivalente podem se tornar diferenciadas em diferentes tipos celulares (ver Capítulo 13).



Produção da vulva de *C. elegans*. A. O tecido definitivo diferenciado. **B.** Método de diferenciação. As células têm um início completamente equivalente. Uma célula-âncora por trás das células equivalentes envia um sinal para as células mais próximas, que formam a vulva. A célula vulvar primária em seguida envia um sinal lateral para suas vizinhas, evitando que elas se tornem células primárias, mesmo que também tenham recebido o sinal da célula-âncora.

O comportamento também tem sido assunto de dissecção genética. *C. elegans* oferece uma vantagem, pois nematódeos com comportamento defeituoso ainda conseguem sobreviver e se reproduzir. Os sistemas nervoso e muscular do verme foram geneticamente dissecados, possibilitando que os comportamentos fossem ligados a genes específicos.

Outra área de contribuição

- Sinalização célula–célula.

Drosophila melanogaster

Organismo-chave para estudar:

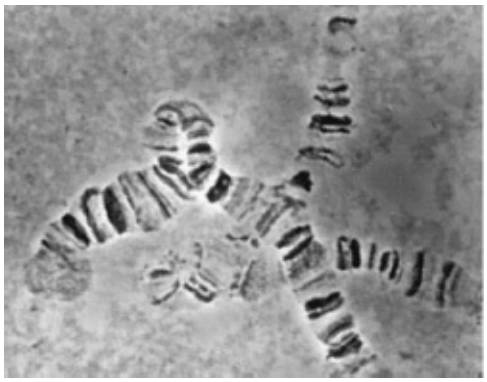
- Genética da transmissão
- Citogenética
- Desenvolvimento
- Genética de populações
- Evolução.

"Estatística vital" genética

Tamanho do genoma:	180 Mb
Cromossomos:	Diploide, 3 autossomos, X e Y ($2n = 8$)
Número de genes:	13.000
Porcentagem com homólogos humanos:	Aproximadamente 50%
Tamanho médio dos genes:	3 kb, 4 éxons/gene
Transpósons:	Elementos *P*, entre outros
Genoma sequenciado em:	2000

A mosca-das-frutas, *Drosophila melanogaster* (livremente traduzida como "apreciadora parda de néctar"), foi um dos primeiros organismos-modelo a ser

utilizado em genética. Ela foi escolhida, em parte, por ser facilmente obtida em frutas maduras, apresentar um ciclo de vida curto do tipo diploide e ser simples de criar e cruzar em potes ou frascos contendo uma camada de alimento. A análise genética inicial demonstrou que os seus mecanismos de herança apresentam fortes semelhanças com aqueles de outros eucariotos, destacando o seu papel como um organismo-modelo. A sua popularidade como organismo-modelo entrou em declínio durante os anos em que *E. coli*, leveduras e outros microrganismos estavam sendo desenvolvidos como ferramentas moleculares. Entretanto, a *Drosophila* vivenciou um renascimento em virtude de ser muito adequada para o estudo da base genética do desenvolvimento, uma das questões centrais da biologia. A importância da *Drosophila* como um modelo para a genética humana é demonstrada pela descoberta de que aproximadamente 60% dos genes considerados causadores de doenças em humanos, bem como 70% dos genes de câncer, apresentam correspondentes na *Drosophila*.



Cromossomos politênicos. (William M. Gelbart, Harvard University.)

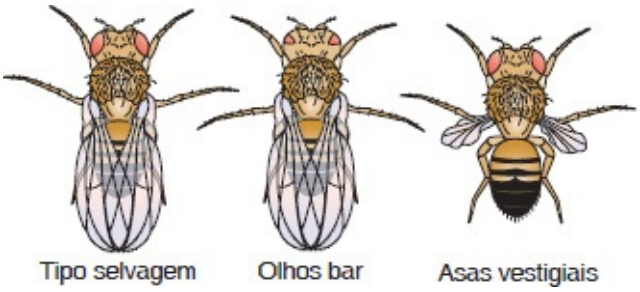
Características especiais

Drosophila entrou em voga como um organismo experimental no início do século 20, em virtude de características comuns à maior parte dos organismos-modelo. Ela é pequena (3 mm de comprimento), simples de criar (originalmente, em garrafas de leite), de reprodução rápida (apenas 12 dias do ovo ao adulto) e de fácil obtenção (apenas deixe alguma fruta apodrecendo). Ela comprovou reunir facilmente uma grande variedade de alelos mutantes interessantes, que foram

utilizados para determinar as regras básicas da genética da transmissão. Os primeiros pesquisadores também se aproveitaram de uma característica única da mosca-das-frutas: os cromossomos politênicos (ver Capítulo 17). Nas glândulas salivares e em outros determinados tecidos, esses "cromossomos gigantes" são produzidos por múltiplas rodadas de replicação do DNA sem segregação cromossômica. Cada cromossomo politênico demonstra um padrão de bandeamento único, que proporciona aos geneticistas pontos de referência que podem ser utilizados para correlacionar mapas com base na recombinação com cromossomos reais. O impulso proporcionado por esses avanços iniciais, juntamente com a grande quantidade de conhecimento acumulado a respeito do organismo, tornou a *Drosophila* um modelo genético atraente.

Análise genética

Cruzamentos em *Drosophila* podem ser realizados de modo consideravelmente fácil. Os genitores podem ser selvagens, estoques mutantes obtidos de centros de estoques ou novas linhagens mutantes.



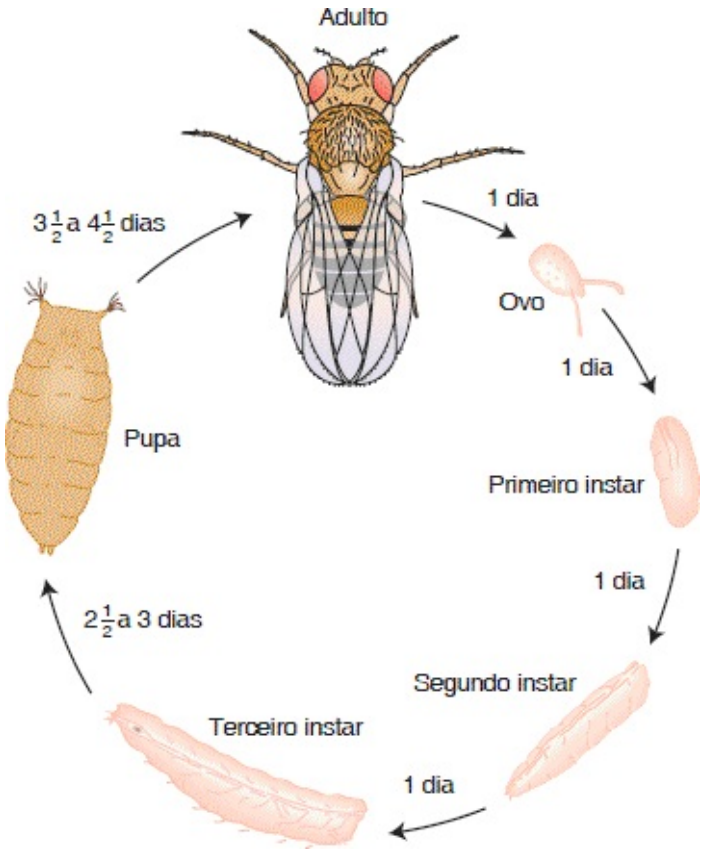
Dois mutantes morfológicos de *Drosophila*, com o tipo selvagem para a comparação.

Ciclo de vida

A *Drosophila* apresenta um ciclo de vida diploide curto, que a torna adequada para a análise genética. Após a eclosão do ovo, a mosca se desenvolve durante diversos estágios de larva e um estágio de pupa antes de surgir como adulta, que logo se torna sexualmente madura. O sexo é determinado pelos

cromossomos sexuais X e Y (XX é fêmea, XY é macho), muito embora, contrariamente aos seres humanos, o número de cromossomos X em relação ao número de autossomos determine o sexo (ver Capítulo 2).

Duração total do ciclo de vida: 12 dias do ovo ao adulto.



Para realizar um cruzamento, machos e fêmeas são colocados em conjunto em um frasco e as fêmeas botam ovos no alimento semissólido que recobre o fundo do frasco. Após surgir das pupas, a progênie pode ser anestesiada para possibilitar a contagem dos membros de classes fenotípicas e para distinguir os machos e as fêmeas (por meio de seus padrões de listras abdominais diferentes). Entretanto, tendo em vista que a progênie feminina permanece virgem durante poucas horas após seu surgimento das pupas, ela deve ser isolada imediatamente se for ser utilizada em cruzamentos controlados. Os cruzamentos destinados à construção de combinações genéticas específicas devem ser cuidadosamente

planejados, tendo em vista que o *crossing over* não ocorre em machos de *Drosophila*. Portanto, em machos, os alelos ligados não se recombinarão para ajudar na criação de novas combinações.

Para a obtenção de novas mutações recessivas, programas de cruzamento especiais (dos quais o protótipo é o teste CIB de Muller) proporcionam sistemas de triagem convenientes. Nesses testes, moscas mutagenizadas são cruzadas com um estoque que apresenta um cromossomo balanceador (ver Capítulo 17). As mutações recessivas são finalmente trazidas à homozigose por autocruzamento durante uma ou duas gerações, iniciando com moscas individuais da F_1.

Técnicas de modificação genética

Mutagênese-padrão:	
Substâncias químicas (EMS) e radiação	Mutações aleatórias nas linhagens germinativa e somática
Transgênese:	
Mediada por elemento *P*	Inserção aleatória
Nocautes gênicos direcionados:	
Substituição induzida	O alelo ectópico nulo sai e se recombina com o alelo do tipo selvagem

RNAi	Mimetiza o nocaute direcionado

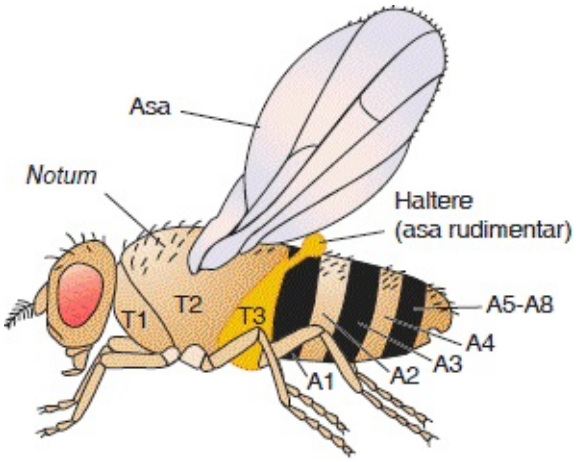
Engenharia genética

Transgênese. A construção de moscas transgênicas requer o auxílio de um transpóson de *Drosophila* denominado elemento *P*. Geneticistas constroem um vetor que carreia um transgene flanqueado por repetições do elemento *P*. O transgene vetor, em seguida, é injetado em um zigoto juntamente com um plasmídio auxiliar que contém uma transposase. A transposase possibilita que o transgene salte aleatoriamente para o genoma nas células germinativas do embrião (ver Capítulo 15).

Nocautes direcionados. Os nocautes de genes-alvo podem ser conquistados, primeiramente, com a introdução de um alelo nulo de modo transgênico em uma posição ectópica e, em segundo lugar, com a indução de enzimas especiais que causam a excisão do alelo nulo. O fragmento excisado (que é linear) em seguida encontra e substitui a cópia endógena por meio do *crossing over* homólogo. Entretanto, nocautes funcionais podem ser produzidos de modo mais eficiente por RNAi.

Contribuições principais

Uma grande parte do desenvolvimento inicial da teoria cromossômica da hereditariedade teve por base os resultados de estudos com *Drosophila*. Geneticistas que trabalham com a *Drosophila* fizeram avanços-chave no desenvolvimento de técnicas para o mapeamento gênico, na compreensão da origem e da natureza da mutação gênica, e na documentação da natureza e do comportamento dos rearranjos cromossômicos (ver Capítulo 4).

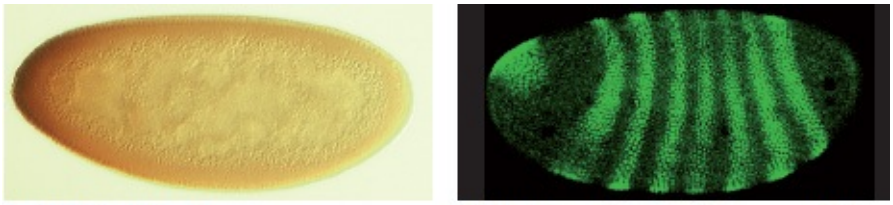


Os segmentos torácico e abdominal normais de *Drosophila*.

Suas descobertas abriram as portas para outros estudos pioneiros:

- Estudos iniciais sobre a cinética da indução de mutação e a estimativa das taxas de mutação foram realizados com a utilização da *Drosophila*. O teste de CIB de Muller e testes semelhantes proporcionaram métodos de triagem convenientes em relação às mutações recessivas
- Rearranjos cromossômicos que movem os genes adjacentes à heterocromatina foram utilizados para descobrir e estudar a variegação por efeito de posição
- Na última parte do século 20, após a identificação de determinadas classes de mutações-chave, tais como as mutações homeóticas e de efeito materno, a *Drosophila* assumiu um papel central na genética do desenvolvimento, um papel que continua atualmente (ver Capítulo 13). As mutações de efeito materno que afetam o desenvolvimento de embriões, por exemplo, têm sido cruciais na elucidação da determinação genética do plano corporal de *Drosophila*; essas mutações são identificadas pela triagem de fenótipos do desenvolvimento anormais nos embriões de uma fêmea específica. Técnicas tais como rastreamento de captura de acentuador possibilitaram a descoberta de novas regiões reguladoras no genoma que afetam o desenvolvimento. Por meio desses métodos e de outros, biólogos que trabalham com *Drosophila* fizeram avanços importantes na compreensão sobre a determinação da segmentação e dos eixos corporais. Alguns dos genes-chave descobertos, tais como os genes homeóticos,

apresentam ampla relevância nos animais em geral.



A. Proteína Bicoid B. Proteína de regra dos pares Hairy

Fotomicrografias demonstrando gradientes de determinantes do plano corporal. A. O mRNA do gene *bcd* está demonstrado localizado na ponta anterior (à esquerda) do embrião. **B.** O mRNA do gene *nos* está localizado posteriormente (ponta à direita do embrião). A distribuição das proteínas codificadas por estes e outros genes determina o eixo corporal. (*A.* Cortesia de Ruth Lehman; *B.* Cortesia de James Langeland.)

Outras áreas de contribuição

- Genética de populações
- Genética evolutiva
- Genética comportamental.

Mus musculus

Organismo-chave para estudar:

- Doenças humanas
- Mutações
- Desenvolvimento
- Cor da pelagem
- Imunologia.

"Estatística vital" genética

Tamanho do genoma:	2.600 Mb

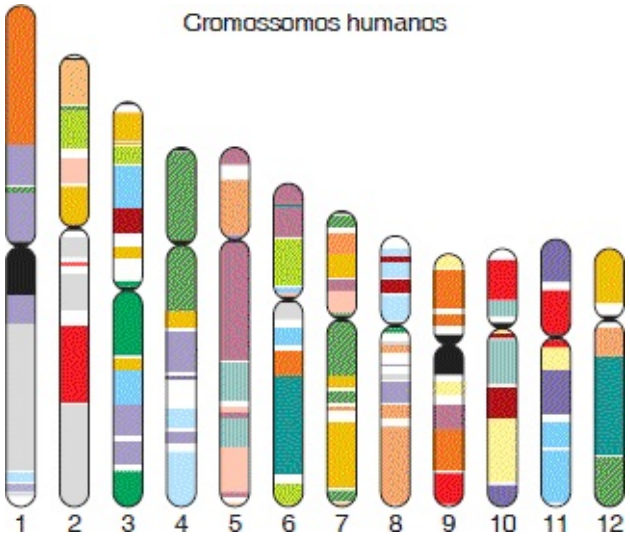
Cromossomos:	19 autossomos, X e Y ($2n = 40$)
Número de genes:	30.000
Porcentagem com homólogos humanos:	99%
Tamanho médio dos genes:	40 kb, 8,3 éxons/gene
Transpósons:	Fonte de 38% do genoma
Genoma sequenciado em:	2002

Tendo em vista que os seres humanos e a maior parte dos animais domesticados são mamíferos, a genética dos mamíferos é de grande interesse. Entretanto, os mamíferos não são ideais para a genética: eles são de tamanho relativamente grande em comparação aos outros organismos-modelo; ocupam, por isso, instalações grandes e dispendiosas; seus ciclos de vida são longos; e seus genomas são grandes e complexos. Entretanto, em comparação aos outros mamíferos, os camundongos (*Mus musculus*) são relativamente pequenos, apresentam ciclos de vida curtos e são facilmente obtidos, o que os torna uma excelente escolha para um modelo em mamíferos. Além disso, os camundongos tiveram uma vantagem na genética, tendo em vista que “apreciadores” de camundongos já haviam desenvolvido muitas linhagens interessantes e diferentes desses animais, que proporcionaram uma fonte de variantes para a análise genética. Pesquisas sobre a genética mendeliana de camundongos começaram no início do século 20.

Camundongo adulto e sua ninhada. (Anthony Griffiths.)

Características especiais

Os camundongos não são exatamente humanos pequenos e peludos, mas a sua constituição genética é notavelmente semelhante à nossa. Entre os organismos-modelo, o camundongo é aquele cujo genoma se assemelha de modo mais próximo ao genoma humano. O genoma do camundongo é aproximadamente 14% menor do que o dos seres humanos (o genoma humano tem 3.000 Mb), mas apresenta aproximadamente o mesmo número de genes (as estimativas atuais são de um pouco menos de 30.000). Surpreendentemente, 99% dos genes do camundongo aparentam apresentar homólogos em seres humanos. Além disso, uma grande proporção do genoma é sintênica com o dos seres humanos; ou seja, existem grandes blocos que contêm os mesmos genes nas mesmas posições relativas (ver Capítulo 14). Tais semelhanças genéticas são a chave para o sucesso do camundongo como um organismo-modelo; elas possibilitam que os camundongos sejam tratados como "substitutos" para seus correspondentes humanos de muitas maneiras. Os possíveis mutágenos e carcinógenos suspeitos de causar danos para os seres humanos, por exemplo, são testados em camundongos, e modelos em camundongos são essenciais no estudo de uma ampla variedade de doenças genéticas humanas.



Um mapa de sintenia entre camundongo-ser humano de 12 cromossomos do genoma humano. A codificação colorida é utilizada para ilustrar as correspondências regionais de cada bloco do genoma humano com as seções correspondentes do genoma do camundongo. Cada cor representa um cromossomo diferente de camundongo.

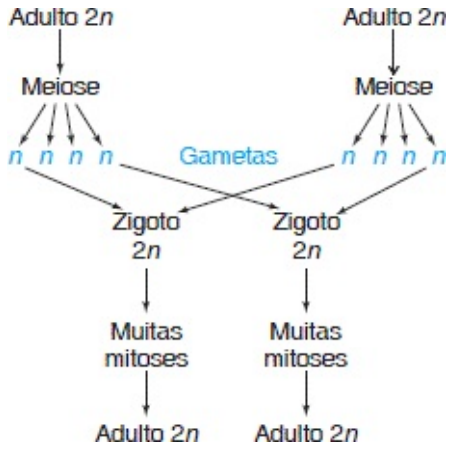
Análise genética

Camundongos mutantes e do "tipo selvagem" (mesmo que não verdadeiramente retirados do ambiente selvagem) são fáceis de ser obtidos: eles podem ser solicitados de grandes centros de estoques, que fornecem camundongos adequados para cruzamentos e diversos outros tipos de experimentos. Muitas dessas linhagens são derivadas de camundongos criados nos séculos passados. Os cruzamentos controlados podem ser realizados simplesmente por meio do pareamento de um macho com uma fêmea não prenhe. Na maior parte dos casos, os genótipos parentais podem ser fornecidos pelo macho ou pela fêmea.

Ciclo de vida

Camundongos apresentam um ciclo de vida diploide familiar, com um sistema de determinação sexual XY semelhante àquele dos seres humanos. As ninhadas têm de 5 a 10 filhotes; entretanto, a fecundidade das fêmeas declina após aproximadamente 9 meses e, assim, elas raramente têm mais de cinco ninhadas.

Duração total do ciclo de vida: 10 semanas desde o nascimento até parir os filhotes, na maior parte das linhagens de laboratório.



A maior parte das estimativas-padrão das taxas de mutação em mamíferos (incluindo seres humanos) tem por base medições em camundongos. Na verdade, os camundongos proporcionam o teste final dos agentes suspeitos de causar mutações em seres humanos. As taxas de mutação na linhagem germinativa são medidas com a utilização do teste *locus* específico: indução de mutagênese nas gônadas +/+, cruzamento com *m/m* (*m* é uma mutação recessiva conhecida no *locus* em estudo), e procura por progênie *m*/m* (*m** é uma nova mutação). O procedimento é repetido em sete *loci* de amostra. A medida das taxas de mutação somática utiliza uma configuração semelhante, mas o mutágeno é injetado no feto. Camundongos têm sido utilizados extensivamente para estudar o tipo de mutação somática que dá origem ao câncer.

Técnicas de modificação genética

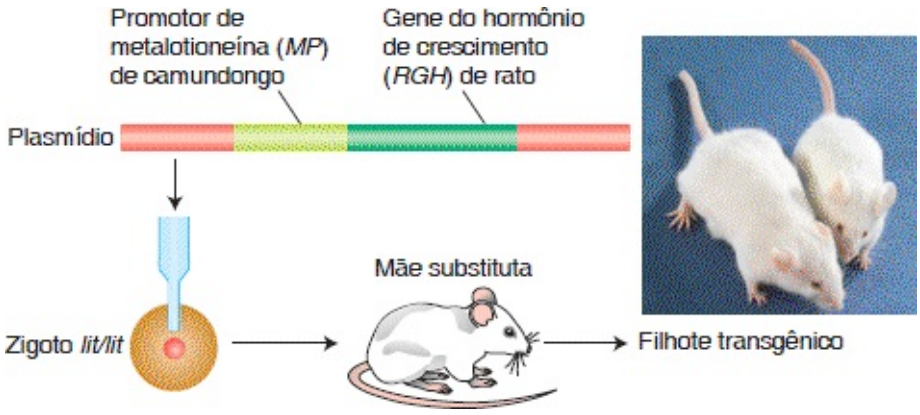
Mutagênese-padrão:

Substâncias químicas e radiação	Mutações nas linhagens germinativa e somática

Transgênese:	
Injeção de transgene no zigoto	Inserção aleatória e homóloga
Captação de transgene por células-tronco	Inserção aleatória e homóloga
Nocautes gênicos direcionados:	
Captação de transgene nulo por células-tronco	Seleção das células-tronco de nocaute direcionado

Engenharia genética

Transgênese. A criação de camundongos transgênicos é direta, mas requer a cuidadosa manipulação de um zigoto (ver Capítulo 10). Primeiramente, o DNA genômico do camundongo é clonado em *E. coli* com a utilização de vetores bacterianos ou fagos. Em seguida, o DNA é injetado em um zigoto, onde ele se integra em locais ectópicos (aleatórios) no genoma ou, menos comumente, no *locus* normal. A atividade da proteína transgênica pode ser monitorada por meio da fusão do transgene com um gene repórter, tal como *GFP*, antes que o gene seja injetado. Com a utilização de um método semelhante, as células somáticas de camundongos também podem ser modificadas por meio da inserção de um transgene: fragmentos específicos de DNA são inseridos em células somáticas individuais e essas células, por sua vez, são inseridas nos embriões de camundongos.



Produção de um camundongo transgênico. O transgene, um gene do hormônio de crescimento de rato, unido a um promotor de camundongo, é injetado em um zigoto de camundongo homozigoto para nanismo (*lit/lit*). (Foto: R. L. Brinster, School of Veterinary Medicine, University of Pennsylvania.)

Nocautes direcionados. Nocautes de genes específicos para a dissecção genética podem ser obtidos por meio da introdução de um transgene que contenha um alelo defeituoso e dois marcadores de resistência a fármacos em uma célula-tronco embrionária do tipo selvagem (ver Capítulo 10). Os marcadores são utilizados para selecionar células transformantes específicas nas quais o alelo defeituoso tenha substituído o alelo do tipo selvagem homólogo. As células transgênicas em seguida são introduzidas em embriões de camundongo. Um método semelhante pode ser utilizado para substituir os alelos do tipo selvagem por um transgene funcional (terapia gênica).

Produção de um nocaute gênico. Um gene de resistência a fármacos (neo^R) é inserido no transgene, tanto para atuar como um marcador quanto para romper o gene, produzindo um nocaute. (O gene *tk* é um segundo marcador.) O construto transgênico em seguida é injetado em células embrionárias de camundongo.

Contribuições principais

Inicialmente na carreira do camundongo como um organismo-modelo, geneticistas utilizavam os camundongos para elucidar os genes que controlam a cor e o padrão da pelagem, fornecendo um modelo em relação a todos os mamíferos que contêm pelos, incluindo gatos, cães, cavalos e gado (ver Capítulo 6). Mais recentemente, estudos da genética de camundongo realizaram uma diversidade de contribuições com impacto direto na saúde humana:

- Uma grande proporção de doenças genéticas humanas apresenta um correspondente em camundongos – denominado um “camundongo-modelo” – útil para estudos experimentais
- Os camundongos atuam como modelos em relação aos mecanismos de mutações em mamíferos
- Estudos sobre os mecanismos genéticos do câncer são realizados em camundongos
- Muitos possíveis carcinógenos são testados em camundongos
- Camundongos têm sido modelos importantes para o estudo da genética do desenvolvimento em mamíferos (ver Capítulo 13). Por exemplo, eles proporcionam um sistema-modelo para o estudo de genes que afetam as fendas labial e palatina, um distúrbio comum do desenvolvimento humano
- Linhagens celulares híbridas, resultantes da fusão dos genomas de camundongos e seres humanos, desempenharam um papel importante na atribuição de genes humanos a cromossomos humanos específicos. Existe uma tendência dos cromossomos humanos de serem perdidos dos referidos híbridos e, assim, a perda de cromossomos específicos pode ser correlacionada à perda de alelos humanos específicos.

Outras áreas de contribuição

- Genética do comportamento
- Genética quantitativa
- Os genes do sistema imune.

Apêndice A

Nomenclatura genética

Não existe um conjunto de normas universalmente aceitas para a denominação de genes, alelos, produtos proteicos e fenótipos correlatos. Primeiramente, cada geneticista desenvolveu seus próprios símbolos para o registro do seu trabalho. Posteriormente, grupos de pessoas que trabalhavam com determinado organismo se reuniam e decidiam sobre um conjunto de convenções que todos utilizariam. Tendo em vista que a *Drosophila* foi um dos primeiros organismos a serem extensivamente utilizados pelos geneticistas, a maior parte dos sistemas atuais é uma variante do sistema de *Drosophila*. Entretanto, existem divergências consideráveis. Alguns cientistas atualmente defendem uma padronização desse simbolismo, a qual, entretanto, não foi realizada. De fato, a situação se tornou mais complexa com o advento da tecnologia do DNA. Embora a maior parte dos genes tenha sido denominada anteriormente em relação aos fenótipos produzidos por mutações neles, a nova tecnologia demonstrou a natureza precisa dos produtos de muitos desses genes. Portanto, aparenta ser mais adequado fazer referência a eles de acordo com a sua função celular. Entretanto, as antigas denominações ainda estão na literatura, de modo que muitos genes apresentam dois conjuntos de nomenclaturas em paralelo.

Os exemplos a seguir de modo algum abrangem todos os organismos utilizados em genética, mas a maior parte dos sistemas de nomenclatura segue um desses tipos.

Drosophila melanogaster (inseto)

ry	Gene que, quando mutado, causa olhos rosados
ry^{502}	Alelo mutante recessivo específico, que produz olhos rosados em homozigotos
ry^{+}	Alelo do tipo selvagem de *rosy*
ry	Fenótipo mutante rosy (rosado)
ry^{+}	Fenótipo do tipo selvagem (olhos vermelhos)
RY	Produto proteico do gene *rosy*

XDH	Xantina desidrogenase, uma descrição alternativa do produto proteico do gene *rosy*; denominada de acordo com a enzima que codifica
D	*Dichaete*; um gene que, quando mutado, causa a perda de determinadas cerdas e a manutenção de asas lateralmente nos heterozigotos e causa letalidade nos homozigotos
D^{3}	Alelo mutante específico do gene *Dichaete*
D^{+}	Alelo do tipo selvagem de *Dichaete*
D	Fenótipo mutante Dichaete
D^{+}	Fenótipo do tipo selvagem
D	(Depende do contexto) Produto proteico do gene *Dichaete* (uma proteína de ligação ao DNA)

Neurospora crassa (fungo)

arg	Gene que, quando mutado, causa necessidade de arginina
arg-1	Gene *arg* específico
arg-1	Alelo mutante não especificado do gene *arg*
arg-1 (1)	Alelo mutante específico do gene *arg-1*
arg-1^{+}	Alelo do tipo selvagem
arg-1	Produto proteico do gene *arg-1*$^{+}$
Arg^{+}	Uma linhagem que não necessita de arginina
Arg^{-}	Linhagem que necessita de arginina

Saccharomyces cerevisiae (fungo)

ARG	Gene que, quando mutado, causa necessidade de arginina
ARG1	Gene *ARG* específico
arg1	Alelo mutante não especificado do gene *ARG*
arg1-1	Alelo mutante específico do gene *ARG1*
ARG1^{+}	Alelo do tipo selvagem
ARG1p	Produto proteico do gene *ARG1*$^{+}$
Arg^{+}	Linhagem que não necessita de arginina
Arg^{-}	Linhagem que necessita de arginina

Homo sapiens (mamífero)

ACH	Gene que, quando mutado, causa acondroplasia
ACH1	Alelo mutante (dominância não especificada)
ACH	Produto proteico do gene *ACH*; natureza desconhecida
FGFR3	Denominação recente para o gene da acondroplasia
FGFR^{1} ou *FGFR3**1 ou *FGFR*3 <1>	Alelo mutante do *FGFR3* (dominância não especificada)
Proteína FGFR3	Receptor 3 do fator de crescimento de fibroblasto

Mus musculus (mamífero)

Tyrc	Gene para tirosinase
$+^{Tyrc}$	Alelo do tipo selvagem desse gene
Tyrc^{ch} ou *Tyrc-ch*	Alelo mutante que causa a coloração chinchila
Tyrc	Produto proteico desse gene
+TYRC	Fenótipo do tipo selvagem
TYRCch	Fenótipo chinchila

Escherichia coli (bactéria)

lacZ	Gene em relação à utilização de lactose
lacZ^{+}	Alelo do tipo selvagem
lacZ1	Alelo mutante
LacZ	Produto proteico desse gene
Lac^{+}	Linhagem capaz de utilizar a lactose (fenótipo)
Lac^{-}	Linhagem incapaz de utilizar a lactose (fenótipo)

Arabidopsis thaliana (planta)

YGR	Gene que, quando mutado, produz folhas amarelo-esverdeadas
YGR1	Gene *YGR* específico
YGR1	Alelo do tipo selvagem

ygr1-1	Alelo mutante recessivo específico de *YGR1*
ygr1-2D	Alelo mutante dominante (D) específico de *YGR1*
YGR1	Produto proteico de *YGR1*
Ygr^{-}	Fenótipo amarelo-esverdeado
Ygr+	Fenótipo do tipo selvagem

Apêndice B

Recursos de bioinformática para genética e genômica

“Você certamente encontrará algo se procurar, mas nem sempre é exatamente o que você estava procurando.” – *O Hobbit*, J. R. R. Tolkien

O campo da bioinformática engloba a utilização de ferramentas computadorizadas para refinar complexos conjuntos de dados. Os dados genéticos e genômicos são tão diversos que se tornou um desafio considerável identificar *sites* com autoridade em relação a um tipo específico de informação. Além disso, o cenário dos *softwares* acessíveis pela Internet para a análise dessas informações está constantemente em alteração na medida em que ferramentas novas e mais poderosas são desenvolvidas. Este apêndice destina-se a fornecer *alguns* pontos de início valiosos para a exploração do universo em rápida expansão dos recursos *on-line* para genética e genômica.

1. Busca por sites sobre genética e genômica

Aqui estão listados diversos recursos centrais que contêm grandes listas de *sites* relevantes:

- O periódico científico denominado *Nucleic Acids Research* (*NAR*) publica uma edição especial todo mês de janeiro listando uma ampla diversidade de recursos de bases de dados *on-line* em http://nar.oupjournals.org/
- A Virtual Library tem subdivisões de organismos-modelo e genética com ricos arranjos de recursos de Internet em http://ceolas.org/VL/mo/
- O National Human Genome Research Institute (NHGRI) mantém uma lista de *sites* de genoma em http://www.nhgri.nih.gov/10000375/
- O Department of Energy (DOE) mantém um *site* do Human Genome Project em http://genomicscience.energy.gov/
- O SwissProt mantém a página de *links* da Amos, em http://expasy.ch/alinks.html.

2. Bases de dados gerais

Bases de dados de sequências de ácidos nucleicos e proteínas. Por um acordo internacional, três grupos colaboram para hospedar as sequências de DNA e mRNA primárias de todas as espécies: o National Center for Biotechnology Information (NCBI) hospeda o GenBank, o European Bioinformatics Institute (EBI) hospeda a biblioteca de dados do European Molecular Biology Laboratory (EMBL) e o National Institute of Genetics, no Japão, hospeda o DNA DataBase of Japan (DDBJ).

Registros da sequência primária do DNA, denominados acessos, são apresentados por grupos de pesquisa individuais. Além de proporcionar o acesso a esses registros de sequências de DNA, esses *sites* fornecem muitos outros conjuntos de dados. Por exemplo, o NCBI também hospeda o RefSeq, uma síntese resumida de informações sobre as sequências de DNA de genomas totalmente sequenciados e dos produtos gênicos codificados por essas sequências.

Muitas outras características importantes podem ser encontradas nos *sites* do NCBI, EBI e da DDBJ. As *home pages* e alguns outros *sites* importantes são:

- **NCBI**: http://www.ncbi.nlm.nih.gov/
- **NCBI-Genomas**: http://www.ncbi.nlm.nih.gov/Genomes/index.html
- **NCBI-RefSeq**: http://www.ncbi.nlm.nih.gov/refseq
- **The UCSC Genome Bioinformatics site**: http://genome.ucsc.edu/.

Esse *site* excepcional contém referência sobre sequências e montagens de rascunhos relativas a uma grande coleção de genomas, além de uma diversidade de ferramentas para explorá-los. O *Genome Browser* foca nos cromossomos, demonstrando o trabalho de pesquisadores mundiais. O *Gene Sorter* demonstra expressão, homologia e outras informações sobre os grupos de genes que podem relacionar-se de muitos modos. *Blat* é uma ferramenta de alinhamento que mapeia rapidamente sequências no genoma. O *Table Browser* proporciona o acesso às bases de dados subjacentes:

- **EBI**: http://www.ncbi.nlm.nih.gov/
- **DDBJ**: http://www.ddbj.nig.ac.jp/.

A dura realidade é que, com tanta informação biológica, o objetivo de tornar esses recursos *on-line* "transparentes" para o usuário não é totalmente alcançado. Portanto, a exploração desses *sites* envolverá a familiarização do usuário com o conteúdo de cada um e a exploração de alguns dos modos pelos quais cada *site* ajuda o usuário a focar nos seus questionamentos, de maneira a obter a(s) resposta(s) correta(s). Como exemplo do poder desses *sites*, considere pesquisar sobre uma sequência de nucleotídios no NCBI. As bases de dados em geral armazenam as informações em pastas separadas, denominadas "campos". Ao utilizar questionamentos que limitam a pesquisa para o campo apropriado, pode ser indagada uma questão mais direcionada. Com a utilização da opção "Limits", uma frase de questionamento pode ser utilizada para identificar ou localizar uma dada espécie, um tipo de sequência (genômica ou de mRNA), um símbolo gênico ou qualquer outra pesquisa de dados. Os mecanismos de consulta normalmente apresentam capacidade de unir múltiplos questionamentos. Por exemplo: recuperar todos os registros de sequências de DNA da espécie *Caenorhabditis elegans* E que foram publicados após 1º de janeiro de 2000. Utilizando a opção "History", os resultados de múltiplos questionamentos podem ser reunidos, de modo que apenas aqueles acertos comuns a múltiplos questionamentos serão recuperados. Por meio da adequada utilização das opções de questionamento disponíveis em um *site*, uma grande parte de falso-positivos pode ser eliminada de modo computadorizado, sem descartar quaisquer dos acertos relevantes.

Tendo em vista que as previsões de sequências proteicas são uma parte natural da análise das sequências de DNA e mRNA, esses mesmos *sites* proporcionam o acesso a uma diversidade de bases de dados de proteínas. Uma base de dados de proteínas importante é SwissProt/TrEMBL. As sequências TrEMBL são previstas automaticamente a partir de sequências de DNA e/ou mRNA. As sequências SwissProt contam com curadoria, o que significa que cientistas especialistas revisam os resultados da análise computadorizada e tomam decisões fundamentadas a respeito dos resultados a ser aceitos ou

rejeitados. Além dos registros de sequências primárias de proteínas, SwissProt também oferece bases de dados de domínios proteicos e assinaturas de proteínas (cadeias de sequências de aminoácidos que são característicos de proteínas de um tipo em particular). A *home page* de SwissProt é http://www.ebi.ac.uk/swissprot/.

Bases de dados de domínios proteicos. Acredita-se que as unidades funcionais existentes nas proteínas sejam regiões de dobramento local denominadas domínios. A previsão de domínios em proteínas recentemente descobertas é um modo de adivinhar a sua função. Surgiram diversas bases de dados de domínios proteicos que preveem os domínios de modos um pouco diferentes. Algumas das bases de dados de domínios individuais são Pfam, PROSITE, PRINTS, SMART, ProDom, TIGRFAMs, BLOCKS e CDD. A InterPro possibilita a consulta a múltiplas bases de dados de domínios proteicos simultaneamente e apresenta os resultados combinados. Os *sites* em relação a algumas bases de dados de domínios são:

- **InterPro**: http://www.ebi.ac.uk/interpro/
- **Pfam**: http://www.sanger.ac.uk/Software/Pfam/index.shtml
- **PROSITE**: http://www.expasy.ch/prosite
- **PRINTS**: http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/
- **SMART**: http://www.smart.embl-heidelberg.de/
- **ProDom**: http://www.prodom.prabi.fr/
- **TIGRFAMs**: http://www.tigr.org/TIGRFAMs/
- **BLOCKS**: http://www.blocks.fhcrc.org/
- **CDD**: http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml.

Bases de dados de estruturas proteicas. A representação das estruturas proteicas tridimensionais se tornou um aspecto importante da análise molecular global. Bases de dados de estruturas tridimensionais estão disponíveis a partir dos principais *sites* de bases de dados de sequências de DNA/proteína e de bases de dados de estruturas proteicas independentes, notavelmente a Protein DataBase

(PDB). O NCBI possui um aplicativo denominado Cn3D que auxilia na visualização dos dados da PDB:

- **PDB**: http://www.rcsb.org/pdb/
- **Cn3D**: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml.

3. Bases de dados especializados

Bases de dados genéticos de organismos específicos. Com a finalidade de reunir algumas classes de informação genética e genômica, especialmente informações fenotípicas, é necessário o conhecimento especializado de uma espécie em particular. Portanto, surgiram os MOD (bases de dados de organismos-modelo; do inglês, *model organism databases*) para preencher esse papel em relação aos principais sistemas genéticos. Essas incluem bases de dados de *Saccharomyces cerevisiae* (SGD), *Caenorhabditis elegans* (WormBase), *Drosophila melanogaster* (FlyBase), do peixe-zebra *Danio rerio* (ZFIN), do camundongo *Mus musculus* (MGI), do rato *Rattus norvegicus* (RGD), de *Zea mays* (MaizeGDB) e *Arabidopsis thaliana* (TAIR). As *home pages* desses MOD podem ser encontradas em:

- **SGD**: http://genome-www.stanford.edu/Saccharomyces/
- **WormBase**: http://www.wormbase.org/
- **FlyBase**: http://flybase.org/
- **ZFIN**: http://zfin.org/
- **MGI**: http://www.informatics.jax.org/
- **RGD**: http://rgd.mcw.edu/
- **MaizeGDB**: http://www.maizegbd.org
- **TAIR**: http://www.arabidopsis.org/.

Bases de dados de genética e genômica humanas. Em virtude da importância da genética humana na clínica, bem como na pesquisa básica, surgiu um conjunto diverso de bases de dados de genética humana. Esse conjunto inclui uma base de dados de doenças genéticas humanas denominada Online Mendelian Inheritance in Man (OMIM), uma base de dados com breve descrição dos genes humanos denominada GeneCards, uma compilação de todas as mutações conhecidas nos genes humanos denominada Human Gene Mutation Database (HGMD), uma base de dados do mapa de sequências atual do genoma humano denominada Golden

Path e alguns *links* para bases de dados de doenças genéticas humanas:

- **OMIM**: http://www.omim.org
- **GeneCards**: http://genecards.org
- **HGMD**: http://www.hgmd.org
- **Golden Path**: http://genome.ucsc.edu/
- **Online Genetic Support Groups**: http://mostgene.org/genetics/genetic-support-groups-directory/
- **Genetic Disease Information**: http://www.geneticalliance.org/diseaseinfo/search.html
- **The Encyclopedia of Genome Elements (ENCODE)**: http://genome.ucsc.edu/ENCODE
- **Cancer Genome Project**: http://cancergenome.nih.gov/.

Bases de dados do Projeto Genoma. Os projetos genoma individuais também têm *sites*, nos quais demonstram os seus resultados, com frequência incluindo informações que não aparecem em qualquer outro *site* no mundo. Os maiores centros de genoma com financiamento público incluem:

- **Whitehead Institute/MIT Center for Genome Research**: http://www-genome.wi.mit.edu/
- **Washington University School of Medicine Genome Sequencing Center**: http://genome.wustl.edu/
- **Baylor College of Medicine Human Genome Sequencing Center**: http://www.hgsc.bcm.tmc.edu
- **Sanger Institute**: http://www.sanger.ac.uk/
- **DOE Joint Genomics Institute**: http://www.jgi.doe.gov/
- **International HapMap Project**: http://hapmap.ncbi.nlm.nih.gov/.

4. Relações dos genes nas bases de dados e entre elas

Os produtos gênicos podem relacionar-se por compartilharem uma origem evolutiva comum, uma função comum ou por participarem da mesma via.

BLAST | Identificação de similaridades na sequência. As evidências em relação a uma origem evolutiva comum advêm da identificação de similaridades entre duas ou mais sequências. Uma das ferramentas mais importantes para a identificação de tais similaridades é a BLAST (Basic Local Alignment Search Tool), que foi desenvolvida pelo NCBI. A BLAST é realmente uma série de programas e bases de dados relacionados, na qual as correspondências locais entre trechos longos de sequências podem ser identificadas e classificadas. A consulta a sequências de DNA ou proteínas similares por meio da BLAST é uma das primeiras coisas que um pesquisador faz com um gene recém-sequenciado. Diferentes bases de dados de sequências podem ser acessadas e organizadas pelo tipo de sequência (genoma de referência, atualizações recentes, EST etc.) e uma espécie ou grupo taxonômico em particular pode ser especificado. Uma rotina da BLAST correlaciona uma sequência indagada de nucleotídios traduzida em todos os seis quadros com uma base de dados de sequência de proteínas. Outra correlaciona uma sequência indagada de proteína com a tradução de seis quadros de uma base de dados de sequências de nucleotídios. Outras rotinas da BLAST são customizadas para identificar correspondências de padrões de sequências curtas ou alinhamentos pareados, para triar segmentos de DNA do tamanho do genoma e assim por diante, e podem ser acessadas por meio da mesma página:

- **NCBI-BLAST**: http://www.ncbi.nlm.nih.gov/BLAST/.

Bases de dados de ontologia de função. Outra abordagem para o desenvolvimento de relações entre os produtos gênicos é por meio da atribuição de papéis funcionais a esses produtos com base na evidência experimental ou previsão. A apresentação de um modo comum para descrever esses papéis, independentemente do sistema experimental é, então, de grande importância. Um grupo de cientistas de diferentes bases de dados está trabalhando para

desenvolver um conjunto comum de termos dispostos hierarquicamente – uma ontologia – em relação à *função* (evento bioquímico), ao *processo* (evento celular para o qual uma proteína contribui) e à *localização subcelular* (localização de um produto em uma célula) como um modo de descrever as atividades de um produto gênico. Essa ontologia em particular é denominada Ontologia Gênica (GO) e muitas diferentes bases de dados de produtos gênicos atualmente incorporam os termos da GO. Uma descrição total pode ser encontrada em:

- http://www.geneontology.org/.

Bases de dados de vias. Ainda outro modo de relacionar os produtos uns aos outros é por meio da sua atribuição a etapas nas vias bioquímicas ou celulares. Diagramas de vias podem ser utilizados como modos organizados de apresentar as relações desses produtos entre si. Algumas das tentativas mais avançadas de produzir as referidas bases de dados de vias incluem a Kyoto Encyclopedia of Genes and Genomes (KEGG) e a Signal Transduction Database (TRANSPATH):

- **KEGG**: http://www.genome.ad.jp/kegg
- **TRANSPATH**: http://genexplain.com/transpath-1.

Glossário

A *Ver* **adenina; adenosina.**

abrigo seguro Sítio no genoma no qual é improvável que a inserção de um elemento de transposição cause uma mutação, prevenindo, assim, o dano ao hospedeiro.

ação gênica Interação de alelos em um *locus*.

ação gênica aditiva Quando o valor de traço em relação à classe heterozigota em um QTL é exatamente intermediário entre os valores de traço em relação às duas classes homozigotas.

ação gênica dominante Situação em que o valor do traço em relação à classe heterozigota em um QTL é igual ao valor do traço em relação a uma das duas classes homozigotas.

acentuador (*enhancer*) Conjunto de proteínas reguladoras composto por fatores de transcrição que se ligam às sequências reguladoras de *ação cis* no DNA.

acentuassomo Montagem macromolecular responsável pela interação de elementos acentuadores e as regiões promotoras dos genes.

ácido desoxirribonucleico *Ver* **DNA**.

ácido ribonucleico *Ver* **RNA**.

adaptabilidade alélica Adaptabilidade média dos indivíduos que carreiam um alelo em particular. É expressa pela equação:

$$W_A = pW_{A/A} + qW_{A/a}$$

adaptabilidade darwiniana Probabilidade relativa de sobrevivência e reprodução de um genótipo.

adaptabilidade média Média da adaptabilidade de todos os membros de uma população.

adaptação No sentido evolutivo, alguma característica herdada do fenótipo de um indivíduo que melhora as suas chances de sobrevivência e reprodução no ambiente existente.

adenina (A) Base purina que pareia com a timina na dupla-hélice do DNA.

adenosina (A) Nucleosídio que contém a adenina como base.

adenosina monofosfato cíclico (cAMP) Molécula que contém uma ligação diéster entre os átomos de carbono 3′ e 5′ da parte ribose do nucleotídio. Esse nucleotídio modificado não pode ser incorporado no DNA nem RNA. Ele desempenha um papel-chave como sinal intracelular na regulação de diversos processos.

agente intercalar Mutágeno que consegue se inserir entre as bases empilhadas no centro da dupla-hélice de DNA, causando uma taxa elevada de *mutações indel*.

agrupamento pi Região nos genomas de vertebrados e invertebrados que codifica agrupamentos de piRNA.

alça de deleção Alça formada na meiose por meio do pareamento de um cromossomo normal e um cromossomo que contém uma deleção.

alça de inversão Alça formada por meio do pareamento meiótico de homólogos em um heterozigoto para inversão.

alelo Uma das diferentes formas de um gene que pode existir em um único *locus*.

alelo dominante Alelo que expressa seu efeito fenotípico até mesmo quando heterozigoto com um alelo recessivo; portanto, se *A* for dominante sobre *a*, então *A/A* e *A/a* apresentam o mesmo fenótipo.

alelo fixado Alelo em relação ao qual todos os membros da população em estudo são homozigotos e, assim, não existem outros alelos em relação a esse *locus* na população.

alelo letal Alelo cuja expressão resulta na morte do organismo que o expressa.

alelo neutro Alelo que não apresenta efeito sobre a adaptabilidade dos

indivíduos que o possuem.

alelo nulo Alelo cujo efeito é a ausência de um produto gênico normal no nível molecular ou de uma função normal no nível fenotípico.

alelo pleiotrópico Alelo que afeta diversas propriedades diferentes de um organismo.

alelo recessivo Alelo cujo efeito fenotípico não é expresso em um heterozigoto.

alelos múltiplos Conjunto de formas de um gene, que diferem em sua sequência de DNA, em sua expressão ou em ambos.

alongamento Estágio da transcrição que segue o início e precede o término.

alopoliploide *Ver* **anfidiploide**.

alta frequência de recombinação celular (Hfr) Em *E. coli*, célula que apresenta seu fator de fertilidade integrado no cromossomo bacteriano; uma célula doadora (masculina).

Alu Elemento de transposição curto, que compõe mais de 10% do genoma humano. Os elementos *Alu* são retroelementos que não codificam proteínas e, como tal, são elementos não autônomos.

aminoácido Peptídio; o elemento estrutural básico das proteínas (ou dos polipeptídios).

aminoacil-tRNA sintetase Enzima que une um aminoácido ao tRNA antes de sua utilização na tradução. Existem 20 aminoacil-tRNA diferentes, um para cada aminoácido.

aminoterminal Extremidade de uma proteína que apresenta um grupo amino livre. Uma proteína é sintetizada a partir da extremidade amino na extremidade de 5′ de um mRNA até a extremidade carboxila próxima da extremidade de 3′ do mRNA durante a tradução.

amostra Pequeno grupo de indivíduos ou observações que são representativos de uma população maior, a partir da qual o grupo foi coletado.

amplificação Produção de muitas cópias de DNA a partir de uma região máster do DNA.

análise de heredogramas Dedução da herança monogênica de fenótipos humanos por meio do estudo da progênie de cruzamentos em uma família, que com frequência abrange diversas gerações.

análogo de base Substância química cuja estrutura molecular mimetiza aquela de uma base do DNA; em virtude da similaridade, o análogo pode atuar como um mutágeno.

aneuploide Genoma que apresenta um número de cromossomos que difere do número cromossômico normal para a espécie em um pequeno número de cromossomos.

anfidiploide Alopoliploide; um poliploide formado a partir da união de dois conjuntos cromossômicos separados e da sua subsequente duplicação.

anotação Identificação de todos os elementos funcionais de um genoma em particular.

anticódon Trinca de nucleotídios em uma molécula de tRNA que se alinha com um códon em particular no mRNA sob a influência do ribossomo; o aminoácido carreado pelo tRNA é inserido em uma cadeia proteica crescente.

anticorpo Molécula proteica (imunoglobulina), produzida pelo sistema imune, que reconhece uma substância em particular (antígeno) e se liga a ela.

antígeno nuclear de proliferação celular (PCNA) Parte do replissomo, o PCNA é a versão eucariótica da proteína procariótica grampo-deslizante (*clamp*).

antiparalelo Termo utilizado para descrever as orientações opostas dos dois filamentos de uma dupla-hélice de DNA; a extremidade de 5′ de um filamento se alinha com a extremidade de 3′ do outro filamento.

antiterminador Proteína que promove a continuação da transcrição ao prevenir o término da transcrição em sítios específicos no DNA.

apoptose Vias celulares responsáveis pela morte celular programada.

arcabouço (1) Estrutura central de um cromossomo, à qual o solenoide do DNA está unido como alças; composto amplamente pela topoisomerase. (2) Em projetos genoma, um conjunto ordenado de *contigs*, no qual pode haver intervalos não sequenciados conectados por leituras de sequência de extremidades pareadas.

arquitetura genética Todos os fatores genéticos e ambientais que influenciam um traço.

asco Em um fungo, saco que envolve uma tétrade ou uma óctade de ascósporos.

associação ampla de genoma (GWA) Mapeamento de associação que utiliza *loci* marcadores por todo o genoma.

atenuação Mecanismo regulador no qual o nível de transcrição de um óperon (tal como *trp*) é reduzido quando o produto final de uma via (p. ex., triptofano) é abundante; a etapa regulada ocorre após o início da transcrição.

atenuador Região da sequência de RNA que forma estruturas secundárias alternativas que regulam o nível de transcrição dos óperons atenuados.

ativadora Proteína que, quando ligada a um elemento de DNA regulador de ação cis, tal como um operador ou um acentuador, ativa a transcrição a partir de um promotor adjacente.

autofecundação Fertilizar ovos com espermatozoides do mesmo indivíduo.

autopoliploide Poliploide formado a partir da duplicação de um único genoma.

autorradiograma Padrão de manchas escuras em um filme fotográfico revelado ou em uma emulsão na técnica de autorradiografia.

auxotrófica Linhagem de microrganismos que proliferará apenas quando o meio for suplementado com uma substância específica não necessária para os organismos do tipo selvagem (*comparar com* **prototrófica**).

BAC Cromossomo artificial bacteriano; um plasmídio F modificado para atuar como um vetor de clonagem que consegue carrear grandes insertos.

bactéria lisogênica Célula bacteriana que contém um profago inerte inserido no cromossomo hospedeiro e que é replicada com ele.

bacteriófago (fago) Vírus que infecta bactérias.

balanceador Cromossomo com múltiplas inversões, utilizado para reter as combinações alélicas favoráveis no homólogo não invertido.

balanço gênico Ideia de que um fenótipo normal exige uma proporção relativa de genes de 1:1 no genoma.

biblioteca de cDNA Biblioteca composta por cDNA, que não necessariamente representa todos os mRNA.

biblioteca de DNA-molde Grupo de moléculas de DNA unifilamentar que pode ser amplificado em muitas cópias.

biblioteca genômica Biblioteca que engloba um genoma inteiro.

bioinformática Sistemas de informação e métodos analíticos computadorizados aplicados a problemas biológicos, tais como a análise genômica.

bivalentes Dois cromossomos homólogos pareados na meiose.

bloqueadores do acentuador Elementos reguladores posicionados entre um promotor e um acentuador. A sua presença previne que o promotor seja ativado pelo acentuador.

bolha de transcrição Sítio no qual a dupla-hélice é desenrolada, de modo que a RNA polimerase possa utilizar um dos filamentos de DNA como um molde para a síntese de RNA.

boxe TATA Sequência de DNA observada em muitos genes eucarióticos, que está localizada aproximadamente 30 pb *upstream* do sítio de início da transcrição.

C *Ver* **citidina; citosina**.

caminhada cromossômica Método para a dissecção de grandes segmentos de DNA, no qual um segmento de DNA clonado, normalmente eucariótico, é

utilizado para triar clones de DNA recombinantes do mesmo banco genômico em relação aos outros clones que contêm sequências vizinhas.

caminho adaptativo Via da alteração evolutiva em etapas, que resulta da seleção natural e do acúmulo de uma série de mutações.

cAMP *Ver* **adenosina monofosfato cíclico**.

câncer Classe de doença caracterizada pela proliferação rápida e descontrolada de células em um tecido de um eucarioto multitecidual. Acredita-se que os cânceres em geral sejam doenças genéticas das células somáticas, que surgem por meio de mutações sequenciais que criam oncogenes e inativam os genes supressores de tumor.

cap Estrutura especial, composta por um resíduo de 7-metilguanosina ligado ao transcrito por três grupamentos fosfato, que é adicionada no núcleo à extremidade 5′ do mRNA eucariótico. O *cap* protege o mRNA contra a degradação e é necessário para a tradução do mRNA no citoplasma.

CAP *Ver* **proteína ativadora de catabólito**.

característica Atributo de membros individuais de uma espécie em relação ao qual podem ser definidas diversas diferenças hereditárias.

carga genética Conjunto total de alelos deletérios em um genótipo individual.

cauda de histona Extremidade de uma proteína histona que se projeta a partir do cerne do nucleossomo e que está sujeita à modificação pós-tradução. *Ver também* **código de histona**.

cauda poli(A) Sequência de nucleotídios adenina adicionada ao mRNA após a transcrição.

cDNA *Ver* **DNA complementar**.

célula F^- Em *E. coli*, célula que não apresenta fator de fertilidade; uma célula feminina.

célula F^+ Em *E. coli*, célula que apresenta um fator de fertilidade livre; uma

célula masculina.

centimorgan (cM) *Ver* **unidade de mapa**.

centro decodificador Região na pequena subunidade ribossômica na qual se decide se um aminoacil-tRNA pode se ligar ao sítio A. Essa decisão tem por base a complementaridade entre o anticódon do tRNA e o códon do mRNA.

centro peptidiltransferase Sítio na grande subunidade ribossômica no qual a ligação de dois aminoácidos é catalisada.

centrômero Região especializada do DNA em cada cromossomo eucariótico que atua como um sítio para a ligação de proteínas do cinetócoro.

ChIP *Ver* **imunoprecipitação de cromatina**.

ciclo lisogênico Ciclo de vida de uma bactéria normal quando ela é infectada por um fago λ do tipo selvagem e o genoma é integrado ao cromossomo bacteriano como um profago inerte.

ciclo lítico Ciclo de vida do bacteriófago que leva à lise da célula hospedeira.

citidina (C) Nucleosídio que contém citosina como a sua base.

citosina (C) Base pirimidina que pareia com a guanina.

citotipo M Estoques laboratoriais de *Drosophila melanogaster* com ausência completa do elemento de transposição *P*, que é observado em estoques naturais (citotipo P).

citotipo P Estoques naturais de *Drosophila melanogaster* que contêm de 20 a 50 cópias do elemento *P*. Os estoques laboratoriais não apresentam nenhuma cópia. *Ver* **citotipo M**.

clonagem posicional Identificação das sequências de DNA que codificam um gene de interesse, com base no conhecimento da sua localização genética ou no mapa citogenético.

clone celular Membros de uma colônia que apresentam um ancestral genético único.

clone de DNA Seção do DNA que foi inserida em um vetor, tal como um plasmídio ou um fago, e em seguida foi replicada para produzir muitas cópias.

CNV *Ver* **variação no número de cópias**.

coativador Classe especial de complexo regulador eucariótico que atua como uma ponte para unir as proteínas reguladoras e a RNA polimerase II.

c.o.c. *Ver* **coeficiente de coincidência**.

código de histona Refere-se ao padrão de modificação (p. ex., acetilação, metilação, fosforilação) das caudas de histonas, que podem carrear as informações necessárias para a correta montagem da cromatina.

código degenerado (redundante) Código genético no qual alguns aminoácidos podem ser codificados por mais de um códon cada.

código genético Conjunto de correspondências entre trincas de nucleotídios no RNA e aminoácidos na proteína.

codominância Situação na qual um heterozigoto demonstra igualmente os efeitos fenotípicos de ambos os alelos.

códon Trecho do RNA (três nucleotídios de comprimento) que codifica um único aminoácido.

coeficiente de autocruzamento (F) Probabilidade de que dois alelos em um *locus* em um indivíduo sejam idênticos por descendência.

coeficiente de coincidência (c.o.c.) Proporção do número observado de recombinantes duplos em relação ao número esperado.

coeficiente de correlação Medida estatística de associação que significa a extensão até a qual duas variáveis variam juntas.

coeficiente de seleção (*s*) Perda da adaptabilidade (ou a desvantagem seletiva) de um genótipo em relação a outro genótipo.

cointegrado Produto da fusão de dois elementos de transposição circulares para formar um círculo único e maior na transposição replicativa.

colônia Clone visível de células.

compensação de dose Processo, em organismos que utilizam um mecanismo de determinação sexual cromossômico (tal como XX *versus* XY), que possibilita aos genes estruturais padrão no cromossomo sexual serem expressos nos mesmos níveis em indivíduos dos sexos feminino e masculino, independentemente do número de cromossomos sexuais. Em mamíferos, a compensação de dose opera por meio da manutenção de apenas um único cromossomo X ativo em cada célula; em *Drosophila*, opera por meio da hiperativação do cromossomo X masculino.

complementação Produção do fenótipo do tipo selvagem quando dois genomas haploides totais ou parciais estão unidos na mesma célula.

complementação funcional (resgate de mutante) Utilização de um fragmento clonado de DNA do tipo selvagem para transformar um mutante no tipo selvagem; utilizada na identificação de um clone que contém um gene específico.

complementar (pares de bases) Faz referência ao pareamento específico entre a adenina e a timina, bem como entre a guanina e a citosina.

complexo de pré-iniciação (PIC) Complexo proteico eucariótico muito grande, que abrange a RNA polimerase II e os seis fatores gerais de transcrição (GTF), cada um dos quais é um complexo de multiproteínas.

complexo gênico Grupo de genes adjacentes relacionados funcional e estruturalmente, que surge tipicamente por meio de duplicação gênica no decorrer da evolução.

complexo mediador Complexo proteico que atua como um adaptador, que interage com fatores de transcrição ligados a sítios reguladores e com fatores de iniciação gerais para a transcrição mediada pela RNA polimerase II.

conformação cis Em um heterozigoto que apresenta dois sítios mutantes em um gene ou em um aglomerado gênico, o arranjo A_1A_2/a_1a_2.

conformação trans Em um heterozigoto com dois sítios mutantes em um gene ou em um aglomerado gênico, o arranjo a_1 +/+ a_2.

conjugação União de duas células bacterianas, durante a qual o material cromossômico é transferido da célula doadora para a receptora.

controle negativo Regulação mediada por fatores que bloqueiam ou interrompem a transcrição.

controle positivo Regulação mediada por uma proteína que é necessária para a ativação de uma unidade de transcrição.

"cópia e colagem" Termo descritivo em relação a um mecanismo de transposição no qual um retrotranspóson da classe 1 é copiado do sítio doador e uma cópia de DNA bifilamentar é inserida (colada) em um novo sítio-alvo. *Ver também* "**corte e colagem**".

corpúsculo de Barr Massa de coloração densa que representa um cromossomo X inativado.

correlação Tendência de uma variável a variar proporcionalmente a outra variável, seja de modo positivo ou negativo.

correpressor Repressor que facilita a repressão gênica, mas que por si próprio não é um repressor de ligação ao DNA.

"corte e colagem" Termo descritivo em relação a um mecanismo de transposição no qual um transpóson da classe 2 (DNA) é excisado (cortado) do sítio doador e inserido (colado) em um novo sítio-alvo. *Ver também* "**cópia e colagem**".

cossupressão Fenômeno epigenético por meio do qual um transgene se torna reversivelmente inativado juntamente com a cópia do gene no cromossomo.

cotransdutores Dois alelos doadores que transduzem simultaneamente uma célula bacteriana; a sua frequência é utilizada como uma medida da proximidade dos genes doadores no mapa cromossômico.

covariância Medida estatística da extensão até a qual duas variáveis alteram juntas. É utilizada no cômputo do coeficiente de correlação entre duas variáveis.

cpDNA DNA de cloroplasto.

cromátide Uma das duas réplicas lado a lado produzidas pela divisão cromossômica.

cromatina Substância dos cromossomos; atualmente sabe-se que inclui o DNA e as proteínas cromossômicas.

cromossomo Arranjo linear, de uma extremidade até a outra, de genes e outro DNA, por vezes com proteínas e RNA associados.

cromossomo acêntrico Cromossomo que não apresenta centrômero.

cromossomo artificial bacteriano *Ver* **BAC**.

cromossomo dicêntrico Cromossomo com dois centrômeros.

cromossomo politênico Cromossomo gigante em tecidos específicos de alguns insetos, produzido por meio de um processo endomitótico no qual os múltiplos conjuntos de DNA permanecem ligados em um número haploide de cromossomos.

cromossomo sexual Cromossomo cuja presença ou ausência está correlacionada ao sexo do portador; um cromossomo que desempenha um papel na determinação do sexo.

cromossomo X Um de um par de cromossomos sexuais, distinto do cromossomo Y.

cromossomo Y Um de um par de cromossomos sexuais, distinto do cromossomo X.

cromossomos homeólogos Cromossomos parcialmente homólogos, que normalmente indicam alguma homologia ancestral comum.

cromossomos homólogos Cromossomos que pareiam entre si na meiose, ou cromossomos em diferentes espécies que retiveram a maior parte dos mesmos genes durante a sua evolução a partir de um ancestral comum.

crossing over Troca de partes cromossômicas correspondente entre homólogos por meio de quebra e reunião.

crRNA RNA transcrito a partir dos *loci* CRISPR que orientam um complexo

proteico a degradar o ácido nucleico viral invasor complementar.

cruzamento Cruzamento deliberado de dois tipos parentais de organismos na análise genética.

cruzamento di-híbrido Cruzamento entre dois indivíduos identicamente heterozigotos em dois *loci* – por exemplo, *AB/ab* × *AB/ab*.

cruzamento interrompido Técnica utilizada para mapear genes bacterianos por meio da determinação da sequência na qual os genes do doador entram nas células receptoras.

cruzamento mono-híbrido Cruzamento entre dois indivíduos identicamente heterozigotos em um par de genes – por exemplo, *A/a* × *A/a*.

cruzamento não preferencial *Ver* **cruzamento preferencial negativo**.

cruzamento preferencial negativo Cruzamento preferencial entre parceiros fenotipicamente não semelhantes.

cruzamento preferencial positivo Situação na qual fenótipos semelhantes cruzam mais comumente do que o esperado ao acaso.

cruzamento-teste Cruzamento de um organismo de genótipo desconhecido ou de um heterozigoto (ou um heterozigoto múltiplo) com um testador.

cruzamento-teste de três pontos (cruzamento-teste de três fatores) Cruzamento-teste no qual um genitor apresenta três pares de genes heterozigotos.

CTD *Ver* **domínio carboxi-terminal**.

deleção Remoção de um segmento cromossômico de um conjunto cromossômico.

deleção intragênica Deleção dentro de um gene.

deleção multigênica Deleção de diversos genes adjacentes.

depressão por endocruzamento Redução no vigor e no sucesso reprodutivo decorrente do endocruzamento.

deriva *Ver* **deriva genética aleatória**.

deriva genética aleatória Alterações na frequência dos alelos que resultam em virtude de os genes que aparecem na descendência não serem uma amostragem perfeitamente representativa dos genes parentais.

desacetilase de histona Atividade enzimática que remove um grupo acetil da cauda de histona, o que promove a repressão da transcrição gênica.

descoberta de gene Processo por meio do qual geneticistas encontram um conjunto de genes que afetam algum processo biológico de interesse por meio dos padrões de herança monogênica de seus alelos mutantes, ou por meio de análise genômica.

desequilíbrio de ligação (LD) Desvio, nas frequências de diferentes haplótipos em uma população, das frequências esperadas se os alelos nos *loci* que definem os haplótipos estiverem aleatoriamente associados.

desoxirribose Açúcar pentose no arcabouço do DNA.

desvio Diferença do valor do traço em um indivíduo do valor médio do traço na população.

desvio genético Segmento de DNA regulador e proteína(s) reguladoras(s) que se ligam a ele, que controlam o estado da transcrição de um gene ou de um conjunto de genes.

desvio padrão Raiz quadrada da variância.

díade Par de cromátides-irmãs unidas no centrômero, como na primeira divisão da meiose.

Dicer Complexo proteico que reconhece longas moléculas de RNA bifilamentar e as cliva em siRNA bifilamentares. Dicer desempenha um papel-chave no RNA de interferência.

diferencial de seleção (*S*) Diferença entre a média de uma população e a média dos membros individuais selecionados para serem os genitores da próxima geração.

di-híbrido Heterozigoto duplo, tal como *A/a* × *B/b*.

dimorfismo Polimorfismo com apenas duas formas.

diploide Célula que apresenta dois conjuntos cromossômicos ou organismo que apresenta dois conjuntos cromossômicos em cada uma de suas células.

diploide parcial *Ver* **merozigoto**.

direcionamento Característica de determinados elementos de transposição, que facilita a sua inserção em regiões do genoma nas quais provavelmente eles não se inserem em um gene, causando uma mutação.

disgenesia híbrida Síndrome que inclui esterilidade, mutação, quebra cromossômica e recombinação masculina na progênie híbrida de cruzamentos entre determinados isolados laboratoriais e naturais de *Drosophila*.

dissecção genética Utilização de recombinação e mutação para juntar os diversos componentes de uma determinada função biológica.

dissômico Haploide anormal que carreia duas cópias de um cromossomo.

distribuição de Poisson Distribuição matemática que fornece a probabilidade de observação de diversos números de um evento em particular em uma amostra, quando a probabilidade média de um evento em qualquer estudo é muito pequena.

distribuição normal Distribuição contínua definida pela função de densidade normal, com uma média e um desvio padrão especificados, que demonstra as frequências esperadas em relação a diferentes valores de uma variável aleatória (a "curva sinusoidal").

diversidade gênica (*GD*) Probabilidade de que dois alelos coletados aleatoriamente do *pool* gênico sejam diferentes.

diversidade nucleotídica Heterozigosidade ou diversidade genética média em todos os sítios de nucleotídios em um gene ou em qualquer outro trecho de DNA.

DNA (ácido desoxirribonucleico) Cadeia de nucleotídios ligados (que apresentam desoxirribose como seu açúcar). Duas cadeias em uma forma de

dupla-hélice são a substância fundamental da qual os genes são compostos.

DNA complementar (cDNA) DNA transcrito a partir de um modelo de RNA mensageiro por meio da ação da enzima transcriptase reversa.

DNA do cloroplasto (cpDNA) Pequeno componente genômico observado nos cloroplastos de plantas, relacionado à fotossíntese e a outras funções que ocorrem na organela.

DNA doador Qualquer DNA utilizado na clonagem ou na transformação mediada por DNA.

DNA hemimetilado Sequência de DNA com um filamento metilado e um filamento não metilado.

DNA heterodúplex DNA no qual existe um par ou mais pares de nucleotídios malpareados em um gene em estudo.

DNA heteroplex DNA no qual existe um par de nucleotídios malpareado em um gene em estudo.

DNA ligase Enzima importante na replicação e no reparo do DNA, que sela o arcabouço do DNA ao catalisar a formação de ligações fosfodiéster.

DNA mitocondrial (mtDNA) Subconjunto do genoma observado na mitocôndria, especializado no fornecimento de algumas das funções da organela.

DNA recombinante Nova sequência de DNA formada por meio da combinação de duas moléculas de DNA não homólogas.

doadora Célula bacteriana utilizada em estudos de transmissão de DNA unidirecional para outras células; exemplos são Hfr na conjugação e fonte de fago na transdução.

dominância completa Descreve um alelo que se expressa do mesmo modo em uma única cópia (heterozigoto) ou em cópia dupla (homozitogo).

dominância incompleta Situação na qual um heterozigoto demonstra um fenótipo quantitativamente (mas não exatamente) intermediário entre os fenótipos

homozigotos correspondentes. (Intermediário exato significa a ausência de dominância.)

dominância parcial Ação gênica sob a qual o fenótipo dos heterozigotos é intermediário entre os dois homozigotos, porém mais semelhante àquele de um dos homozigotos do que do outro.

dominância total *Ver* **dominância completa**.

dominante Fenótipo exibido por um heterozigoto.

domínio Região de uma proteína associada a uma função em particular. Algumas proteínas contêm mais de um domínio.

domínio carboxi-terminal (CTD) Cauda proteica da subunidade β da RNA polimerase II; ela coordena o processamento dos pré-mRNA eucarióticos, incluindo a adição do *cap*, a recomposição e o término.

domínio de ativação Parte de um fator de transcrição necessário para a ativação da transcrição do gene-alvo; pode se ligar aos componentes do maquinário de transcrição ou recrutar proteínas que modificam a estrutura da cromatina, ou ambos.

domínio de ligação ao DNA Sítio em uma proteína de ligação ao DNA que interage diretamente com sequências de DNA específicas.

downstream Modo de descrever a localização relativa de um sítio em uma molécula de DNA ou RNA. Um sítio *downstream* está localizado mais próximo da extremidade 3′ de uma unidade de transcrição.

dsRNA *Ver* **RNA bifilamentar**.

dupla-hélice Estrutura do DNA proposta pela primeira vez por James Watson e Francis Crick, com duas hélices entrelaçadas e unidas por ligações de hidrogênio entre as bases pareadas.

duplicação Mais de uma cópia de um segmento cromossômico em particular em um conjunto cromossômico.

duplicação do sítio-alvo Sequência de DNA de repetição direta curta (tipicamente de 2 a 10 pb de comprimento) adjacente às extremidades de um elemento de transposição que foi gerado durante a integração do elemento no cromossomo hospedeiro.

duplicação em *tandem* Segmentos cromossômicos idênticos adjacentes.

duplicação gênica Duplicação de genes ou segmentos de DNA por meio da replicação errônea do DNA.

duplicação insercional Duplicação na qual a cópia extra não está adjacente à cópia normal.

duplicação segmentar Presença de duas ou mais grandes repetições não em *tandem*.

efeito aditivo Metade da diferença entre a média dos valores fenotípicos em relação às classes genotípicas homozigotas em um QTL.

efeito da dosagem gênica (1) Proporcionalidade da expressão de alguma função biológica e do número de cópias de um alelo presente na célula. (2) Alteração no fenótipo causada por um número anormal de alelos do tipo selvagem (observada nas mutações cromossômicas).

efeito de dominância Diferença entre o valor do traço em relação à classe heterozigota em um QTL e o ponto médio entre os valores do traço das duas classes homozigotas.

efeito de posição Descreve uma situação na qual a influência fenotípica de um gene é alterada por modificações na posição do gene no genoma.

efeito fundador Diferença aleatória na frequência de um alelo ou um genótipo em uma nova colônia, em comparação à população parental que resulta de um pequeno número de fundadores.

efeito sinergístico Característica das proteínas reguladoras eucarióticas, em relação à qual a ativação da transcrição mediada pela interação de diversas

proteínas é maior do que a soma dos efeitos das proteínas considerados individualmente.

efetor alostérico Pequena molécula que se liga a um sítio alostérico.

elemento *Ac* Ver **elemento *Ativador***.

elemento acentuador Sequência reguladora de *ação cis* que pode elevar os níveis de transcrição de um promotor adjacente. Muitos acentuadores tecido-específicos podem determinar os padrões espaciais de expressão gênica em eucariotos superiores. Os acentuadores podem atuar em promotores ao longo de muitas dezenas de quilobases de DNA e podem estar a 5′ ou 3′ dos promotores que regulam.

elemento *Ativador* (*Ac*) Elemento de transposição de DNA da classe 2, assim denominado pela sua descobridora, Barbara McClintock, em virtude de ser necessário para ativar a quebra cromossômica no *locus* de *dissociação* (*Ds*).

elemento autônomo Elemento de transposição que codifica a(s) proteína(s) – por exemplo, transposase ou transcriptase reversa – necessária(s) para a sua transposição e para a transposição de elementos não autônomos na mesma família.

elemento curto intercalado (SINE) Tipo de elemento de transposição da classe 1 que não codifica transcriptase reversa, mas que se acredita utilizar a transcriptase reversa codificada pelos LINE. *Ver também* ***Alu***.

elemento da sequência de inserção (IS) Parte móvel do DNA bacteriano (diversas centenas de pares de nucleotídios de comprimento), capaz de inativar um gene no qual se insere.

elemento de ação cis Sítio em uma molécula de DNA (ou RNA) que atua como um sítio de ligação para uma proteína de ligação ao DNA (ou RNA) sequência-específica. O termo *de ação cis* indica que a ligação proteica a esse sítio afeta apenas as sequências de DNA (ou RNA) próximas na mesma molécula.

elemento de *Dissociação* (*Ds*) Elemento de transposição não autônomo,

denominado por Barbara McClintock em virtude de sua capacidade de quebrar o cromossomo 9 de milho, mas apenas na presença de outro elemento, denominado *Ativador* (*Ac*).

elemento de transposição da classe 1 Elemento de transposição que se movimenta por meio de um RNA intermediário. Também denominado elemento de RNA ou retroelemento.

elemento de transposição da classe 2 Elemento de transposição que se movimenta diretamente de um sítio no genoma até outro. Também denominado elemento de DNA.

elemento DNA Elemento de transposição da classe 2, observado em procariotos e em eucariotos, assim denominado em virtude da participação direta do elemento de DNA na transposição.

elemento IS *Ver* **elemento da sequência de inserção**.

elemento longo intercalado (LINE) Tipo de elemento de transposição da classe 1 que codifica uma transcriptase reversa. Os LINE também são denominados retrotranspósons não de LTR.

elemento não autônomo Elemento de transposição que depende dos produtos proteicos de elementos autônomos para a sua mobilidade. *Dissociação* (*Ds*) é um exemplo de um elemento de transposição não autônomo.

elemento *P* Elemento de transposição do DNA em *Drosophila*, que foi utilizado como uma ferramenta para a mutagênese por inserção e para a transformação de linhagem germinativa.

elemento proximal do promotor Série de sítios de ligação de fatores de transcrição localizados próximos do promotor.

elemento tipo *copia* Elemento de transposição (retrotranspóson) de *Drosophila* que é flanqueado por longas repetições terminais e que tipicamente codifica uma transcriptase reversa.

elemento transponível miniatura com repetições invertidas (MITE) Tipo de

transpóson de DNA não autônomo, que pode ser formado por meio da deleção do gene da transposase de um elemento autônomo e alcançar números de cópias muito grandes.

elemento *Ty* Retrotranspóson LTR de levedura; o primeiro isolado de qualquer organismo.

eletroforese em gel Método de separação molecular no qual o DNA, o RNA ou proteínas são separadas em uma matriz de gel de acordo com o tamanho molecular, com a utilização de um campo elétrico para direcionar as moléculas através do gel em um sentido predeterminado.

embrioide Pequena massa em divisão de células monoploides, produzida a partir de uma célula destinada a se tornar uma célula de pólen por meio de sua exposição ao frio.

endocruzamento (*inbreeding*) Cruzamento entre parentes.

endogenoto *Ver* **merozigoto**.

engenharia genética Processo de produção de DNA modificado em um tubo de ensaio e de reintrodução desse DNA em organismos hospedeiros.

enzima de restrição Endonuclease que reconhece sequências de nucleotídios-alvo específicas no DNA e que quebra a cadeia de DNA nesses pontos; é conhecida uma diversidade dessas enzimas e elas são extensivamente utilizadas na engenharia genética.

enzima distributiva Enzima que consegue adicionar apenas um número limitado de nucleotídios antes da liberação do molde de DNA.

enzima processiva Conforme utilizada no Capítulo 7, descreve o comportamento da DNA polimerase III, que pode realizar milhares de rodadas de catálise sem a dissociação de seu substrato (o filamento-modelo de DNA).

epigenética Alterações químicas não genéticas em histonas ou no DNA, que alteram a função gênica sem alterar a sequência do DNA.

epistasia Situação na qual a expressão fenotípica diferencial de um genótipo em

um *locus* depende do genótipo em outro *locus*; uma mutação que exerce a sua expressão enquanto cancela a expressão dos alelos de outro gene.

epistasia de sinal Dependência da vantagem ou desvantagem adaptativa de uma nova mutação nas mutações que foram fixadas anteriormente.

equilíbrio de Hardy-Weinberg Distribuição de frequências estável dos genótipos *A/A*, *A/a* e *a/a*, nas proporções de p^2, 2 *pq* e q^2, respectivamente (em que *p* e *q* são as frequências dos alelos *A* e *a*), que é consequência de um cruzamento aleatório na ausência de mutação, migração, seleção natural ou deriva aleatória.

equilíbrio de ligação Adequação perfeita das frequências de haplótipos em uma população às frequências esperadas se os alelos nos *loci* que definem os haplótipos estiverem aleatoriamente associados.

espécie dioécia Espécie de planta na qual os órgãos masculino e feminino estão em plantas separadas.

EST *Ver* **etiqueta de sequência expressa**.

estrutura populacional Divisão de uma espécie ou de uma população em múltiplas subpopulações geneticamente distintas.

estrutura primária de uma proteína Sequência de aminoácidos na cadeia polipeptídica.

estrutura quaternária de uma proteína Constituição multimérica de uma proteína.

estrutura secundária de uma proteína Arranjo em espiral ou zigue-zague da cadeia polipeptídica.

estrutura terciária de uma proteína Dobramento ou enrolamento em espiral da estrutura secundária para formar uma molécula globular.

estrutura teta (θ) Estrutura intermediária na replicação de um cromossomo bacteriano circular.

estruturas reiteradas de maneira seriada Partes corporais que são membros de

séries repetidas, tais como dígitos, costelas, dentes, membros e segmentos.

etiqueta de sequência expressa (EST) Clone de cDNA em relação ao qual apenas as extremidades 5′ e/ou 3′ foram sequenciadas; utilizada para identificar extremidades de transcritos na análise genômica.

eucromatina Região cromossômica menos condensada, que se acredita conter a maior parte dos genes normalmente funcionais.

euploide Célula que apresenta qualquer número de conjuntos cromossômicos completos ou um organismo individual composto por essas células.

evolução neutra Alterações evolutivas não adaptativas em virtude da deriva genética aleatória.

excisar Descreve o que um elemento de transposição realiza quando deixa um local cromossômico. Também denominado *transpor*.

exconjugante Célula bacteriana feminina que acabou de se conjugar com uma masculina e que contenha um fragmento de DNA masculino.

exogenoto *Ver* **merozigoto**.

éxon Qualquer seção não um íntron da sequência codificadora de um gene; em conjunto, os éxons correspondem ao mRNA que é traduzido em proteína.

expansão de trincas Expansão de uma repetição de 3 pb, a partir de um número de cópias relativamente baixo até um alto número de cópias, que é responsável por uma diversidade de doenças genéticas, tais como a síndrome do X frágil e a doença de Huntington.

experimento de pulso e caça Experimento no qual as células são cultivadas em meio radioativo durante um breve período (o pulso) e em seguida são transferidas para um meio não radioativo durante um período mais longo (a caça).

expressão constitutiva Refere-se aos genes que são expressos continuamente, independentemente das condições biológicas.

expressão gênica Processo por meio do qual a sequência de DNA de um gene é

transcrita em RNA e, em relação aos genes que codificam proteínas, em um polipeptídio.

expressividade Grau até o qual um determinado genótipo é expresso no fenótipo.

extranuclear Refere-se a uma pequena fração especializada dos genomas eucarióticos observada em mitocôndrias ou cloroplastos.

extremidade carboxila Extremidade de uma proteína que apresenta um grupo carboxila livre. A extremidade carboxila é codificada pela extremidade de 3′ do mRNA e é a última parte da proteína a ser sintetizada na tradução.

fago *Ver* **bacteriófago**.

fago temperado Fago que pode se tornar um profago.

fago virulento Fago que não consegue se tornar um profago; a infecção pelo referido fago sempre leva à lise da célula hospedeira.

família de genes Conjunto de genes em um genoma, todos descendentes do mesmo gene ancestral.

fator de ação trans Molécula reguladora difusível (quase sempre uma proteína) que se liga a um elemento de ação cis específico.

fator de fertilidade (fator F) Episssomo bacteriano cuja presença confere capacidade de doador (caráter masculino).

fator de iniciação Proteína necessária para o correto início da tradução.

fator de liberação (RF) Proteína que se liga ao sítio A do ribossomo quando um códon de fim está no mRNA.

fator de transcrição geral (GTF) Complexo proteico eucariótico que não participa na síntese do RNA, mas que se liga à região promotora para atrair e posicionar corretamente a RNA polimerase II para o início da transcrição.

fator F *Ver* **fator de fertilidade**.

fator F′ Fator de fertilidade no qual uma parte do cromossomo bacteriano foi

incorporada.

fatores R Plasmídios que carreiam genes que codificam resistência a diversos antibióticos.

fator sigma (σ) Proteína bacteriana que, como parte da holoenzima RNA polimerase, reconhece as regiões –10 e –35 de promotores bacterianos, posicionando, assim, a holoenzima para iniciar a transcrição corretamente no sítio de início. O fator σ se dissocia da holoenzima antes da síntese do RNA.

fenótipo (1) Forma assumida por alguma característica (ou um grupo de características) em um indivíduo específico. (2) Manifestações externas detectáveis de um genótipo específico.

fenótipo instável Fenótipo caracterizado por reversão frequente, seja de modo somático ou germinativo, ou ambos, em virtude da interação de elementos de transposição com um gene do hospedeiro.

filamento codificador Filamento não molde de uma molécula de DNA, que apresenta a mesma sequência do transcrito de RNA.

filamento contínuo (*leading*) Na replicação do DNA, o filamento que é sintetizado no sentido de 5′ para 3′ por meio da polimerização contínua na ponta 3′ em crescimento.

filamento de RNA antissenso Filamento de RNA que apresenta uma sequência complementar a um filamento de RNA transcrito.

filamento descontínuo (*lagging*) Na replicação do DNA, o filamento que é sintetizado aparentemente no sentido de 3′ para 5′ por meio da ligação de fragmentos curtos sintetizados individualmente no sentido de 5′ para 3′.

filogenia História evolutiva de um grupo.

flutuação genética Alteração na frequência de um alelo em uma população, que resulta de diferenças ao acaso nos números reais da progênie de diferentes genótipos, produzidos por diferentes membros individuais.

fluxo gênico *Ver* **migração**.

forma ceto *Ver* **mudança tautomérica**.

forma enol *Ver* **mudança tautomérica**.

forma imino *Ver* **mudança tautomérica**.

forquilha de replicação Ponto no qual os dois filamentos de DNA são separados para possibilitar a replicação de cada filamento.

fosfato Íon formado por quatro átomos de oxigênio unidos a um átomo de fósforo, ou o grupo químico formado pela união de um íon fosfato a outra espécie química por meio de uma ligação éster.

fosmídio Vetor que consegue carrear um inserto de 35 a 45 kb de DNA exógeno.

FR *Ver* **frequência de recombinantes**.

fragmento acêntrico Fragmento cromossômico que não apresenta centrômero.

fragmento de Okazaki Pequeno segmento de DNA unifilamentar, sintetizado como parte do filamento descontínuo na replicação do DNA.

fragmento de restrição Fragmento de DNA que resulta do corte do DNA com uma enzima de restrição.

frequência alélica Medida do quanto um alelo é comum em uma população; a proporção de todos os alelos daquele gene na população que são desse tipo específico.

frequência de recombinantes (FR) Proporção (ou porcentagem) de células ou indivíduos recombinantes.

frequência genotípica Proporção de indivíduos em uma população que apresentam um genótipo em particular.

função de mapeamento Fórmula que expressa a relação entre a distância em um mapa de ligação e a frequência de recombinantes.

G *Ver* **guanina**.

gargalo Período de uma ou diversas gerações consecutivas de contração no tamanho da população.

GD *Ver* **diversidade gênica**.

gene Unidade física e funcional fundamental da hereditariedade, que carreia a informação de uma geração até a próxima; um segmento de DNA composto por uma região transcrita e uma sequência reguladora que torna a transcrição possível.

gene candidato Gene que, em virtude da sua posição cromossômica ou de alguma outra propriedade, torna-se um candidato a uma função em particular, tal como risco de doença.

gene constitutivo Termo informal em relação a um gene cujo produto é necessário em todas as células e que realiza uma função fisiológica básica.

gene de efeito materno Gene que produz um efeito apenas quando presente na mãe.

gene de polaridade segmentar Em *Drosophila*, membro de uma classe de genes que contribuem para os aspectos finais do estabelecimento do número correto de segmentos. As mutações de polaridade do segmento causam perda ou alteração em uma parte comparável de cada um dos segmentos corporais.

gene de regra dos pares Em *Drosophila*, um membro de uma classe de genes expressos em zigotos que atuam em um estágio intermediário no processo de estabelecimento do número correto de segmentos corporais. As mutações de regra dos pares apresentam metade do número normal de segmentos, em virtude da perda de segmentos alternados.

gene endógeno Gene normalmente presente em um organismo, contrariamente a um gene exógeno de um organismo diferente, que pode ser introduzido por meio de técnicas transgênicas.

gene essencial Gene sem pelo menos uma cópia da qual o organismo morre.

gene *gap* Em *Drosophila*, classe de genes cardinais que são ativados no zigoto

em resposta aos gradientes anteroposteriores das informações posicionais.

gene hemizigoto Gene presente em apenas uma cópia em um organismo diploide – por exemplo, um gene ligado ao X em um mamífero do sexo masculino.

gene repórter Gene cuja expressão fenotípica é de fácil monitoramento; utilizado para estudar as atividades do promotor e do acentuador tecido-específico em transgenes.

gene *SRY* Gene da masculinidade, que está localizado no cromossomo Y.

gene supressor de tumor Gene que codifica uma proteína que suprime a formação de tumor. Acredita-se que os alelos do tipo selvagem dos genes supressores de tumor atuem como reguladores negativos da proliferação celular.

genes controlados de maneira coordenada Genes cujos produtos são simultaneamente ativados ou reprimidos em paralelo.

genes *Hox* Membros dessa classe de genes são os genes homeóticos que contêm homeoboxes agrupados, que controlam a identidade das partes corporais ao longo do eixo anteroposterior da maior parte dos animais bilaterais.

genética (1) Estudo dos genes. (2) Estudo da hereditariedade.

genética de populações Estudo da variação genética nas populações e das alterações ao longo do tempo na quantidade ou na padronização daquela variação, que resulta de mutação, migração, recombinação, deriva genética aleatória, seleção natural e sistemas de cruzamento.

genética direta Abordagem clássica da análise genética, na qual os genes são identificados primeiramente por meio dos alelos mutantes e dos fenótipos mutantes e, em seguida, clonados e submetidos à análise molecular.

genética molecular Estudo dos processos moleculares subjacentes à estrutura e à função gênica.

genética quantitativa Subcampo da genética que estuda a herança de traços complexos ou quantitativos.

genética reversa Procedimento experimental, que tem início com um segmento de DNA clonado ou uma sequência proteica e que o utiliza (por meio de mutagênese direcionada) para introduzir mutações programadas de volta no genoma para investigar a função.

genoma Complemento inteiro de material genético em um conjunto cromossômico.

genômica Clonagem e caracterização molecular de genomas inteiros.

genômica comparativa Análise das relações das sequências do genoma de duas ou mais espécies.

genômica funcional Estudo dos padrões de expressão de transcritos e proteínas e de interações moleculares no nível do genoma como um todo.

genômica pessoal Análise do genoma de um indivíduo para compreender melhor a sua ancestralidade ou a base genética de traços fenotípicos, tal como o seu risco de desenvolver uma doença.

genótipo Composição alélica de um indivíduo ou de uma célula – seja do genoma inteiro ou, mais comumente, de um determinado gene ou de um conjunto de genes.

geração F_1 Primeira geração filial, produzida pelo cruzamento de duas linhagens parentais.

geração F_2 Segunda geração filial, produzida por autofecundação ou intercruzamento da geração F_1.

geração parental Duas linhagens ou dois organismos que constituem o início de um experimento de cruzamento genético; sua progênie constitui a geração F_1.

GGR *Ver* **reparo genômico global**.

grampo β (beta *clamp*) Proteína que circunda o DNA como uma rosquinha, mantendo a enzima DNA pol III unida à molécula de DNA na forquilha de replicação.

grupo externo Taxa de saída de um grupo de organismos, entre os quais as

relações evolutivas estão sendo determinadas.

guanina (G) Base purina que pareia com a citosina.

GWA *Ver* **associação ampla de genoma**.

H *Ver* **heterozigosidade**.

haploide Célula que apresenta um conjunto cromossômico ou um organismo composto por essas células.

haplossuficiente Descreve um gene que, em uma célula diploide, consegue promover a função do tipo selvagem em apenas uma cópia (dose).

haplótipo Tipo (ou forma) de um segmento haploide de um cromossomo, conforme definido pelos alelos presentes nos *loci* daquele segmento.

HapMap Mapa de haplótipos do genoma inteiro.

helicase Enzima que quebra as ligações de hidrogênio no DNA e que desenrola o DNA durante a movimentação da forquilha de replicação.

herança complexa Tipo de herança exibida pelos traços afetados por uma mistura de fatores genéticos e ambientais. Os traços contínuos, tais como a altura, tipicamente apresentam herança complexa.

herança epigenética Modificações herdáveis na função gênica, não em virtude de alterações na sequência de bases do DNA do organismo. Exemplos de herança epigenética são a paramutação, a inativação do cromossomo X e o *imprinting* parental.

herança materna Tipo de herança uniparental, no qual toda a progênie apresenta o genótipo e o fenótipo do genitor que atua como o sexo feminino.

herança simples Tipo de herança no qual apenas um gene (ou alguns poucos genes) está(ão) envolvido(s) e no qual o ambiente apresenta pouco ou nenhum efeito sobre o fenótipo; os traços categóricos com frequência exibem herança simples.

herança uniparental Padrão de herança no qual a progênie apresenta o genótipo e o fenótipo de apenas um genitor; por exemplo, herança de genomas mitocondriais.

herdabilidade no sentido amplo (H^2) Proporção da variância fenotípica total no nível populacional que é contribuição da variância genética.

herdabilidade no sentido restrito (h^2) Proporção de variância fenotípica que pode ser atribuída à variância genética aditiva.

heterocárion Cultura de células composta por dois tipos nucleares diferentes em um citoplasma comum.

heterocromatina Regiões cromossômicas condensadas de coloração densa, que se acredita serem, em sua maior parte, geneticamente inertes.

heterocromatina constitutiva Regiões cromossômicas de cromatina permanentemente condensada, normalmente ao redor dos telômeros e dos centrômeros.

heterozigosidade Medida da variação genética em uma população; em relação a um *locus*, citada como a frequência de heterozigotos para esse *locus*.

heterozigoto Organismo que apresenta um par de genes em heterozigose.

heterozigoto para inversão Diploide com um homólogo normal e um invertido.

hexaploide Célula que apresenta seis conjuntos cromossômicos, ou um organismo composto pelas referidas células.

Hfr *Ver* **alta frequência de recombinação celular**.

hibridizar (1) Formar um híbrido por meio da realização de um cruzamento. (2) Parear filamentos complementares de ácidos nucleicos de diferentes fontes.

hiperacetilação Superabundância de grupos acetil ligados a determinados aminoácidos das caudas de histonas. A cromatina transcricionalmente ativa normalmente é hiperacetilada.

hipoacetilação Carência de grupos acetil em determinados aminoácidos das

caudas de histonas. A cromatina transcricionalmente inativa normalmente é hipoacetilada.

hipótese multifatorial Hipótese que explica a variação quantitativa ao propor que os traços são controlados por um grande número de genes, cada um com um pequeno efeito sobre o traço.

hipótese nula Em estatística, a hipótese que está sendo testada, que faz uma previsão a respeito dos resultados esperados de um experimento. Se a probabilidade de observação dos resultados sob a hipótese nula for inferior a 0,05, a hipótese nula então é rejeitada.

hipótese um gene–um polipeptídio Hipótese de meados do século 20, que originalmente propunha que cada gene (sequência de nucleotídios) codifica uma sequência polipeptídica; em geral verdadeira, com exceção do RNA funcional não traduzido.

histograma de frequência "Curva de medidas", na qual as frequências de diversas classes arbitrariamente delimitadas são plotadas.

histona Tipo de proteína básica que forma a unidade ao redor da qual o DNA está enrolado nos nucleossomos dos cromossomos eucarióticos.

holoenzima DNA polimerase III (holoenzima DNA pol III) Em *E. coli*, o grande complexo multissubunidades na forquilha de replicação, composto por dois centros catalíticos e muitas proteínas acessórias.

holoenzima pol III *Ver* **holoenzima DNA polimerase III**.

holoenzima polimerase III *Ver* **holoenzima DNA polimerase III**.

holoenzima RNA polimerase Complexo de multissubunidades bacteriano composto por quatro subunidades do cerne da enzima, mais o fator σ.

homeoboxe (boxe homeótico) Família de sequências de DNA de 180 pb consideravelmente semelhantes, que codifica uma sequência polipeptídica denominada homeodomínio, uma sequência de ligação ao DNA de sequência específica. Embora o homeoboxe tenha sido descoberto pela primeira vez em

todos os genes homeóticos, atualmente sabe-se que codifica um motivo de ligação ao DNA muito mais difundido.

homeodomínio Família de sequências altamente conservadas, de 60 aminoácidos de comprimento e encontradas em um grande número de fatores de transcrição, que podem formar estruturas de hélice-giro-hélice e ligar-se ao DNA de modo sequência-específico.

homólogo Membro de um par de cromossomos homólogos.

homozigose Refere-se ao estado de carrear um par de alelos idênticos em um *locus*.

homozigoto Organismo com característica de homozigose.

homozigoto dominante Refere-se a um genótipo tal como *A/A*.

homozigoto recessivo Refere-se a um genótipo tal como *a/a*.

IBD *Ver* **idêntico por ascendência**.

idêntico por ascendência (IBD) Quando duas cópias de um gene em um indivíduo remontam à mesma cópia em um ancestral.

ilha CpG Dinucleotídios CG não metilados observados em agrupamentos próximos dos promotores gênicos.

impressão de DNA Padrão de bandeamento autorradiográfico produzido quando o DNA é digerido com uma enzima de restrição que corta uma família de VNTR (número variável de repetições em *tandem*) e um *Southern blot* do gel de eletroforese é hibridizado com uma sonda VNTR específica. Contrariamente às impressões digitais verdadeiras, esses padrões não são únicos de cada organismo individual.

***imprinting* genômico** Fenômeno no qual um gene herdado de um dos genitores não é expresso, muito embora ambas as cópias do gene sejam funcionais. Os genes imprintados são metilados e inativados na formação dos gametas masculinos ou femininos.

***imprinting* materno** Expressão de um gene apenas quando herdado do pai, tendo em vista que a cópia do gene herdada da mãe é inativa em virtude da metilação no período de formação do gameta.

***imprinting* paterno** Expressão de um gene apenas quando herdado da mãe, tendo em vista que o alelo do gene herdado do pai é inativo em virtude da metilação no período de formação do gameta.

imunoprecipitação da cromatina (ChIP) Utilização de anticorpos para isolar regiões específicas da cromatina e para identificar as regiões do DNA às quais as proteínas reguladoras estão ligadas.

indução (1) Alívio da repressão de um gene ou de um conjunto de genes sob controle negativo. (2) Interação de duas ou mais células ou dois ou mais tecidos, que é necessária para que uma dessas células ou um desses tecidos altere o curso de seu desenvolvimento.

indução zigótica Súbita liberação de um fago lisogênico de um cromossomo Hfr quando o profago entra na célula F^-, seguida pela subsequente lise da célula receptora.

indutor Agente ambiental que aciona a transcrição de um óperon.

infecção dupla (mista) Infecção de uma bactéria por dois fagos geneticamente diferentes.

infecção mista (dupla) Infecção de uma cultura bacteriana com dois fagos de genótipos diferentes.

inferência filogenética Determinação do estado de uma característica ou sentido da alteração em uma característica com base em sua distribuição em uma filogenia de organismos.

informação posicional Processo por meio do qual indicações químicas que estabelecem o destino celular ao longo de um eixo geográfico são estabelecidas em um embrião em desenvolvimento ou em um primórdio tecidual.

iniciação Primeiro estágio da transcrição ou da tradução. A sua principal função

na transcrição é posicionar corretamente a RNA polimerase antes do estágio de alongamento, e na tradução é posicionar corretamente o primeiro aminoacil-tRNA no sítio P.

iniciador tRNA especial que insere o primeiro aminoácido de uma cadeia polipeptídica no sítio P ribossômico no início da tradução. O aminoácido carreado pelo iniciador em bactérias é a *N*-formilmetionina.

integração ectópica Em um organismo transgênico, a inserção de um gene em outro sítio que não o seu *locus* habitual.

interatoma Conjunto total de interações moleculares dentro das células, incluindo em particular interações de proteínas.

interferência Medida da independência dos *crossovers* entre si, calculada por meio da subtração do coeficiente de coincidência de 1.

íntron *Ver* **sequência interveniente**.

íntron autorremovível Primeiro exemplo de RNA catalítico; nesse caso, um íntron que pode ser removido de um transcrito sem o auxílio de uma enzima.

inversão Mutação cromossômica composta pela remoção de um segmento cromossômico, por sua rotação em 180º e sua reinserção no mesmo local.

inversão paracêntrica Inversão que não inclui o centrômero.

inversão pericêntrica Inversão que inclui o centrômero.

isoformas Relacionadas por diferentes proteínas. Podem ser originadas por meio da recomposição alternativa de um gene.

isolador de barreira Elemento de DNA que evita a difusão da heterocromatina ao atuar como um sítio de ligação para as proteínas que mantêm modificações da cromatina eucromática, tais como a acetilação de histonas.

isolamento pela distância Desvio na escolha do parceiro para cruzamento que tem origem a partir da dimensão da distância geográfica entre os indivíduos, fazendo com que os indivíduos estejam mais aptos a cruzar com um membro mais

próximo do que com outro membro de sua espécie mais distante.

junção de pontas não homólogas (NHEJ) Mecanismo utilizado pelos eucariotos para reparar as quebras bifilamentares.

LD *Ver* **desequilíbrio de ligação**.

lei da segregação independente (segunda lei de Mendel) Pares de genes não ligados ou ligados de modo distante, ao segregarem, distribuem-se independentemente na meiose.

lei da segregação igual (primeira lei de Mendel) Produção de números iguais (50%) de cada alelo nos produtos meióticos (p. ex., gametas) de um meiócito heterozigoto.

lei de Hardy-Weinberg Equação utilizada para descrever a relação entre as frequências alélicas e genotípicas em uma população de cruzamento aleatório.

leituras de pontas pareadas Na montagem da sequência genômica pela estratégia de *shotgun*, sequências de DNA que correspondem a ambas as extremidades de um inserto de DNA genômico em um clone recombinante.

lesão espontânea Dano do DNA que ocorre na ausência de exposição a mutágenos; ocorre primariamente em virtude da ação mutagênica dos subprodutos do metabolismo celular.

letal sintético Refere-se a um mutante duplo que é letal, embora as mutações únicas constituintes não sejam.

ligação ao sexo Localização de um gene em um cromossomo sexual.

ligação ao X Padrão de herança dos genes encontrados no cromossomo X, mas não no cromossomo Y.

ligação ao Y Padrão de herança dos genes encontrados no cromossomo Y, mas não no cromossomo X (raro).

ligado Situação na qual dois genes estão mesmo cromossomo, conforme deduzido

por meio das frequências de recombinantes inferiores a 50%.

LINE *Ver* **elemento longo intercalado**.

linhagem pura Reserva composta por indivíduos geneticamente idênticos que foram totalmente autocruzados a partir de genitor(es) comum(ns).

linhagem pura População de indivíduos na qual todos contêm um genótipo homozigoto totalmente idêntico.

linhagem quase isogênica *Ver* **linhagens congênicas**.

linhagens congênicas Linhagens ou estoques de uma espécie que são idênticas em seus genomas, com exceção de uma pequena região de interesse.

lisado População de progênie de fago.

lise Ruptura e a morte de uma célula bacteriana na liberação da progênie de fago.

lisógena *Ver* **bactéria lisogênica**.

lncRNA *Ver* **Longo RNA não codificador**.

***loci* CRISPR** Regiões em cromossomos bacterianos que contêm agrupamentos de repetições palindrômicas curtas regularmente interespaçadas, envolvidas na imunidade a vírus.

locus (plural, ***loci***) *Ver* ***locus* gênico**.

***locus* do traço quantitativo (QTL)** Gene que contribui para a variação fenotípica em um traço que demonstra padrões complexos de herança, tais como altura e peso.

***locus* gênico** Local específico em um cromossomo no qual um gene está localizado.

longa repetição terminal (LTR) Repetição direta de uma sequência de DNA nas extremidades 5′ e 3′ de retrovírus e retrotranspósons.

longo RNA não codificador (lncRNA) Transcritos não codificadores de proteína que apresentam aproximadamente mais de 200 nucleotídios de comprimento.

LTR *Ver* **longa repetição terminal**.

LTR solo Única cópia de um *LTR*.

mapa cromossômico Representação de todos os cromossomos no genoma como linhas, marcada com as posições dos genes conhecidos a partir de seus fenótipos mutantes, mais marcadores moleculares. Tem por base a análise da frequência de recombinantes.

mapa de associação Método para a localização dos *loci* dos traços quantitativos no genoma com base no desequilíbrio de ligação entre um *locus* marcador e o *locus* do traço quantitativo em uma população de cruzamento aleatório.

mapa de ligação Mapa cromossômico; um mapa abstrato de *loci* cromossômicos, que tem por base as frequências de recombinantes.

mapa de recombinação Mapa cromossômico no qual as posições dos *loci* demonstradas têm por base as frequências de recombinantes.

mapa físico Mapa ordenado e orientado de fragmentos de DNA clonados no genoma.

mapeamento de deleção Utilização de um conjunto de deleções conhecidas para mapear novas mutações recessivas por meio da pseudodominância.

mapeamento de *locus* do traço quantitativo Método para a localização de QTL no genoma e para a caracterização dos efeitos do QTL sobre a variação do traço.

mapeamento fino Encontro da localização genômica de um gene de interesse (ou uma região funcional dentro de um gene) com *loci* marcadores que estão muito estreitamente ligados a ele.

marca epigenética Alteração hereditária, tal como a metilação do DNA ou uma modificação de histona, que deixa a sequência do DNA inalterada.

marcação de transpóson Método utilizado para identificar e isolar um gene hospedeiro por meio da inserção de um elemento de transposição clonado no gene.

marcador genético Alelo utilizado como uma sonda experimental para acompanhar um organismo, um tecido, uma célula, um núcleo, um cromossomo ou um gene individual.

marcador microssatélite Diferença no DNA no mesmo *locus* em dois genomas, que ocorre em virtude de comprimentos de repetições diferentes de um microssatélite.

marcador minissatélite *Locus* heterozigoto que representa um número variável de repetições em *tandem* de uma unidade com comprimento de 15 a 100 nucleotídios.

marcador não selecionado Em um experimento de recombinação bacteriana, alelo marcado na progênie pela frequência da sua cossegregação com um alelo selecionado ligado.

matriz de leitura aberta (ORF) Seção do tamanho de um gene de um segmento de DNA sequenciado, que inicia com um códon de iniciação e que termina com um códon de fim; presume-se que seja a sequência codificadora de um gene.

média Média aritmética.

meio mínimo Meio que contém apenas sais inorgânicos, uma fonte de carbono e água.

meiócito Célula na qual ocorre a meiose.

meiose Duas divisões nucleares sucessivas (com as divisões celulares correspondentes), que produzem gametas (em animais) ou esporos sexuais (em plantas e fungos) que apresentam metade do material genético da célula original.

merozigoto Célula de *E. coli* parcialmente diploide, formada a partir de um cromossomo completo (o endogenoto) mais um fragmento (o exogenoto).

metilação do DNA Adição de grupos metil a resíduos de DNA após a replicação.

microarranjo Conjunto de moléculas de DNA que contêm todos ou a maior parte dos genes em um genoma depositados em um pequeno *chip* de vidro.

microRNA (miRNA) Classe de RNA funcional que regula a quantidade de proteína produzida por um gene eucariótico.

microssatélite *Locus* composto por diversas cópias (repetições) de um motivo de sequência curta (aproximadamente 2 a 6 pb). Diferentes alelos apresentam diferentes números de repetições.

migração Movimentação de indivíduos (ou gametas) entre populações.

miRNA *Ver* **microRNA**.

mistura genética Mistura de genes que resulta quando os indivíduos apresentam ancestrais de mais de uma subpopulação.

MITE *Ver* **elemento transponível miniatura com repetições invertidas**.

mitose Tipo de divisão nuclear (que ocorre na divisão celular), que produz dois núcleos-filhos idênticos ao núcleo parental.

modificação de histona Alteração covalente de um ou mais resíduos de aminoácidos da proteína histona. As modificações incluem acetilação, fosforilação e metilação.

modificação pós-tradução (PTM) Alteração de resíduos de aminoácidos que ocorre após a proteína ter sido traduzida.

modificadora Mutação em um segundo *locus*, que altera o grau de expressão de um gene mutado em um primeiro *locus*.

molde "Molde" molecular, que forma a estrutura ou a sequência de outra molécula; por exemplo, a sequência de nucleotídios do DNA atua como um molde para controlar a sequência de nucleotídios do RNA durante a transcrição.

molécula-filha Um dos dois produtos da replicação do DNA, composto por um filamento-molde e um filamento recém-sintetizado.

mono-híbrido Heterozigoto de *locus* único do tipo *A/a*.

monoploide Célula que apresenta apenas um conjunto cromossômico (normalmente como uma aberração), ou um organismo composto pelas referidas

células.

monossômico Célula ou um organismo que é basicamente diploide, mas que apresenta apenas uma cópia de um tipo cromossômico em particular e que, portanto, apresenta um número cromossômico de $2n + 1$.

montagem de sequência Compilação de milhares ou milhões de leituras de sequências de DNA independentes em um conjunto de *contigs* e andaimes.

morfo Forma de um polimorfismo genético; o morfo pode ser um fenótipo ou uma sequência molecular.

mRNA (RNA mensageiro) Molécula de RNA transcrita a partir do DNA de um gene; uma proteína é traduzida a partir dessa molécula de RNA por meio da ação dos ribossomos.

mtDNA DNA mitocondrial.

mudança tautomérica Isomerização espontânea de uma base nitrogenada de sua forma ceto normal para uma forma alternativa enol (ou imino) de ligação ao hidrogênio.

mutação (1) Processo que produz um gene ou um conjunto cromossômico que difere daquele do tipo selvagem. (2) Gene ou conjunto cromossômico que resulta de um referido processo.

mutação constitutiva Alteração em uma sequência de DNA que faz com que um gene que está reprimido por vezes seja expresso continuamente, ou "constitutivamente".

mutação cromossômica Qualquer tipo de alteração na estrutura ou no número de cromossomos.

mutação de ponto Mutação que altera a posição de uma única base em uma molécula de DNA ao convertê-la em uma base diferente ou por meio da inserção/deleção de uma única base em uma molécula de DNA.

mutação de sentido trocado Substituição de um par de nucleotídios em uma região codificadora de proteínas, que leva à substituição de um aminoácido por

outro aminoácido.

mutação espontânea Mutação que ocorre na ausência de exposição a mutágenos.

mutação hipomórfica (*leaky*) Mutação que confere um fenótipo mutante, mas que ainda retém um nível baixo, porém detectável, de função do tipo selvagem.

mutação indel Mutação na qual um ou mais pares de nucleotídios são adicionados ou deletados.

mutação induzida Mutação que tem origem por meio da ação de um agente que aumenta a taxa na qual as mutações ocorrem.

mutação na matriz de leitura (*frameshift mutation*) Inserção ou deleção de um ou mais pares de nucleotídios, que causa uma ruptura na matriz de leitura da tradução.

mutação negativa dominante Alelo mutante que, em dose única (um heterozigoto), anula a função gênica por meio de um efeito espoliador sobre a proteína.

mutação nula Mutação que resulta em ausência completa da função do gene.

mutação sem sentido Substituição de um par de nucleotídios em uma região codificadora de proteína, que altera um códon em relação a um aminoácido para um códon de término (sem sentido).

mutação sinônima Mutação que altera um códon para um aminoácido em outro códon para o mesmo aminoácido. Também denominada *mutação silenciosa*.

mutação termossensível Mutação condicional que produz o fenótipo mutante em uma faixa de temperaturas e o fenótipo do tipo selvagem em outra faixa de temperaturas.

mutagênese Experimento no qual organismos experimentais são tratados com um mutágeno e a sua progênie é examinada em relação a fenótipos mutantes específicos.

mutagênese por inserção Situação na qual uma mutação tem origem a partir da

interrupção de um gene por um DNA exógeno, tal como a partir de um construto transgênico ou de um elemento de transposição.

mutágeno Agente capaz de aumentar a taxa de mutação.

mutante Organismo ou uma célula que carreia uma mutação.

mutante duplo Genótipo com alelos mutantes de dois genes diferentes.

mutante resistente Mutante que consegue crescer em um ambiente normalmente tóxico.

NAHR *Ver* **recombinação homóloga não alélica**.

não disjunção Falha de homólogos (na meiose) ou de cromátides irmãs (na mitose) em se separar adequadamente para polos opostos.

ncRNA *Ver* **RNA não codificador de proteína**.

neofuncionalização Evolução de uma nova função por parte de um gene.

NER *Ver* **sistema de reparo por excisão de nucleotídio**.

NH *Ver* **número de haplótipos**.

NHEJ *Ver* **junção de pontas não homólogas**.

NLS *Ver* **sequência de localização nuclear**.

nocaute gênico Inativação de um gene por meio de uma mutação de ocorrência natural, ou por meio da integração de um fragmento de DNA especialmente modificado e introduzido. Em alguns sistemas, a referida inativação é aleatória, com a utilização de construtos transgênicos que se inserem em muitos locais diferentes no genoma. Em outros sistemas, pode ser realizado de modo direcionado. *Ver também* **nocaute gênico direcionado**.

nocaute gênico direcionado Introdução de uma mutação nula em um gene por meio de uma alteração projetada em uma sequência de DNA clonada, que em seguida é introduzida no genoma por meio de recombinação homóloga e da substituição do alelo normal.

Northern blot Transferência de moléculas de RNA separadas por eletroforese de um gel para uma membrana absorvente, que em seguida é imersa em uma sonda marcada que se ligará ao RNA de interesse.

nucleossomo Unidade básica da estrutura cromossômica eucariótica; uma esfera de oito moléculas de histonas, envolta por duas voltas de DNA.

nucleotídio Molécula composta por uma base nitrogenada, um açúcar e um grupo fosfato; o elemento estrutural básico dos ácidos nucleicos.

nulissômico Refere-se a uma célula ou um organismo com um tipo cromossômico ausente, com um número de cromossomos tal como $n - 1$ ou $2n - 2$.

número de haplótipos (NH) Contagem simples do número de haplótipos em um *locus* em uma população.

número variável de repetições em *tandem* (VNTR) *Locus* cromossômico no qual uma sequência repetida em particular está presente em diferentes números em diferentes indivíduos, ou nos dois homólogos diferentes em um indivíduo diploide.

O *Ver* **origem da replicação**.

óctade Asco que contém oito ascósporos, produzidos em espécies nas quais a tétrade normalmente é submetida a uma divisão mitótica pós-meiótica.

OGM *Ver* **organismo geneticamente modificado**.

oncogene Mutação de ganho de função que contribui para a produção de câncer.

oncoproteína Produto proteico de uma mutação em oncogene.

operador Região do DNA em uma extremidade de um óperon, que atua como o sítio de ligação para uma proteína repressora.

óperon Conjunto de genes estruturais adjacentes, cujo mRNA é sintetizado em um segmento, mais os sinais reguladores adjacentes que afetam a transcrição dos genes estruturais.

ORF *Ver* **matriz de leitura aberta**.

organismo geneticamente modificado (OGM) Termo popular para um organismo transgênico, especialmente aplicado a produtos agrícolas transgênicos.

organismo transgênico Organismo cujo genoma foi modificado por um novo DNA aplicado externamente.

origem (O) *Ver* **origem da replicação**.

origem da replicação (O) Ponto de uma sequência específica no qual a replicação do DNA é iniciada.

ortólogos Genes em diferentes espécies que evoluíram a partir de um gene ancestral comum por especiação.

oscilação (*wobble*) Capacidade de determinadas bases na terceira posição de um anticódon no tRNA de formar ligações de hidrogênio de diversos modos, causando alinhamento com diversos diferentes códons possíveis.

padrão de segregação de primeira divisão (padrão de M_I) Padrão linear de fenótipos de esporos dentro de um asco em relação a um par de alelos em particular, produzido quando os alelos se dirigem para núcleos separados na primeira divisão meiótica, demonstrando que não ocorreu nenhum *crossover* entre o par de alelos e o centrômero.

padrão de segregação de segunda divisão (padrão M_{II}) Padrão de genótipos de ascósporos em relação a um par de genes, que demonstra que os dois alelos separam-se em núcleos diferentes apenas na segunda divisão meiótica, como um resultado de um *crossover* entre aquele par de genes e o seu centrômero; pode ser detectado apenas em um asco linear.

padrão M_I *Ver* **padrão de segregação de primeira divisão**.

padrão M_{II} *Ver* **padrão de segregação de segunda divisão**.

painel de descoberta Grupo de indivíduos utilizados para detectar sítios de nucleotídios variáveis, comparando-se as sequências genômicas parciais desses

indivíduos entre si.

palíndromo de DNA Segmento de DNA no qual ambos os filamentos apresentam a mesma sequência de nucleotídios, porém em orientação antiparalela.

par de genes Duas cópias de um tipo de gene em particular, presente em uma célula diploide (uma em cada conjunto cromossômico).

par de genes heterozigoto Par de genes que apresenta alelos diferentes nos dois conjuntos cromossômicos do indivíduo diploide – por exemplo, *A/a* ou A^1/A^2.

paradoxo de valor *C* Discrepância (ou ausência de correlação) entre o conteúdo de DNA de um organismo e a sua complexidade biológica.

parálogos Genes relacionados pela duplicação gênica em um genoma.

parcimônia Favorecimento da explicação mais simples, que envolve o menor número de alterações evolutivas.

pareamento de filamento dependente de síntese (SDSA) Mecanismo livre de erro para a correção de quebras bifilamentares que ocorrem após a replicação de uma região cromossômica em uma célula em divisão.

partenogênese Produção de descendência por parte de uma fêmea sem a contribuição genética de um macho.

PCNA *Ver* **antígeno nuclear de proliferação celular**.

PCR *Ver* **reação da cadeia de polimerase**.

penetrância Proporção de indivíduos com um genótipo específico, que manifesta aquele genótipo no nível do fenótipo.

pentaploide Organismo com cinco conjuntos de cromossomos.

pequeno RNA de interferência (siRNA) RNA bifilamentares curtos produzidos pela clivagem de RNA bifilamentares longos por Dicer.

pequeno RNA nuclear (snRNA) Qualquer um de diversos RNA curtos observados no núcleo eucariótico, no qual auxiliam em eventos de processamento do RNA.

PEV *Ver* **variegação por efeito de posição**.

PIC *Ver* **complexo de pré-iniciação**.

pirimidina Tipo de base nitrogenada; as bases pirimidinas no DNA são a citosina e a timina.

piRNA *Ver* **RNA de interação *piwi***.

pirossequenciamento Tecnologia de sequenciamento do DNA, que tem por base a geração e a detecção de um grupo pirofosfato liberado de um nucleotídio trifosfato.

placa Área clara em uma camada bacteriana, deixada pela lise das bactérias por infecções progressivas por um fago e seus descendentes.

plaqueamento Difusão das células de um microrganismo (bactéria, fungo) em uma placa de meio nutritivo, para possibilitar que cada célula forme uma colônia visível.

plaqueamento em réplica Em genética microbiana, modo para triar colônias dispostas em uma placa máster, para verificar se elas são mutantes em outros ambientes; é utilizada uma almofada de feltro para transferir as colônias para novas placas.

plasmídio Molécula de DNA extracromossômico que se replica de modo autônomo.

plasmídio F′ *Ver* **fator F′**.

plasmídio R Plasmídio que contém um ou diversos transpósons que contêm genes de resistência.

plasmídio Ti Plasmídio circular de *Agrobacterium tumifaciens* que possibilita que a bactéria infecte células de plantas e produza um tumor (tumor galha da coroa).

poligene (*locus* do traço quantitativo) Gene cujos alelos são capazes de interagir de modo aditivo com alelos em outros *loci* para afetar um fenótipo

(traço) que demonstra uma distribuição contínua.

polimerase de *bypass* (translesão) DNA polimerase que consegue continuar a replicar o DNA passando por um sítio de dano que interromperia a replicação pela polimerase replicativa normal. As polimerases de *bypass* contribuem para um mecanismo de tolerância aos danos, denominado síntese de DNA translesão.

polimerases translesão Família de DNA polimerases que conseguem continuar a replicar o DNA além do sítio de uma lesão que interromperia a replicação pela polimerase replicativa normal. Também conhecidas como *polimerases de bypass*.

polimorfismo Ocorrência, em uma população, de múltiplas formas de um traço ou de múltiplos alelos em um *locus* genético.

polimorfismo de nucleotídio único (SNP) Diferença em um par de nucleotídios em uma determinada localização nos genomas de dois ou mais indivíduos de ocorrência natural.

polimorfismo do comprimento de sequência simples (SSLP) Existência, na população, de indivíduos com diferentes números de cópias de uma sequência de DNA simples curta em um *locus* cromossômico.

polimorfismo do comprimento do fragmento de restrição (RFLP) Diferença na sequência de DNA entre indivíduos ou haplótipos, que é reconhecida como diferentes comprimentos de fragmentos de restrição. Por exemplo, a substituição de um par de nucleotídios pode causar a presença de um sítio de reconhecimento de enzima de restrição em um alelo de um gene e a ausência em outro. Consequentemente, uma sonda para essa região do DNA hibridizará com fragmentos de diferentes tamanhos no DNA digerido por enzima de restrição desses dois alelos.

polipeptídio Cadeia de aminoácidos ligados; uma proteína.

poliploide Célula que apresenta três ou mais conjuntos cromossômicos, ou um organismo composto pelas referidas células.

ponte anafásica Em um cromossomo dicêntrico, o segmento entre os centrômeros

que são puxados para os polos opostos na divisão nuclear.

ponte dicêntrica Em um cromossomo dicêntrico, o segmento entre os centrômeros que é puxado para polos opostos na divisão nuclear.

***pool* gênico** Soma total de todos os alelos nos membros reprodutivos de uma população em um determinado momento.

população (1) Grupo de indivíduos que cruzam entre si para produzir a próxima geração. (2) Grupo de indivíduos a partir do qual uma amostra é coletada.

pré-mRNA *Ver* **transcrito primário**.

primase Enzima que produz *primers* de RNA na replicação do DNA.

primeira geração filial (F_1) Indivíduos da progênie que surge a partir de um cruzamento de duas linhagens diploides homozigotas.

primeira lei de Mendel (lei da segregação igual) Dois membros de um par de genes se segregam na meiose; cada gameta apresenta uma probabilidade igual de obter qualquer membro do par de genes.

primer Oligonucleotídio de RNA ou DNA, que pode atuar como um molde para a síntese do DNA pela DNA polimerase quando ligado a uma molécula de DNA mais longa.

primossomo Complexo proteico na forquilha de replicação, cujo componente central é a primase.

probando Em um heredograma humano, pessoa que primeiramente chamou a atenção do geneticista.

procarioto Organismo composto por uma célula procariótica, tal como uma bactéria ou uma alga azul-esverdeada.

processamento cotranscricional Transcrição e o processamento simultâneos do pré-mRNA eucariótico.

processamento de RNA Termo coletivo em relação às modificações do RNA eucariótico, incluindo a adição do *cap* e a recomposição, que são necessárias

antes que o RNA possa ser transportado para o citoplasma para a tradução.

processamento pós-transdução Modificações de grupos laterais de aminoácidos após uma proteína ter sido traduzida.

produto da meiose Uma das células (normalmente quatro) formadas nas duas divisões meióticas.

produtos de *crossover* Células que são produto de meiose com cromossomos que participaram de um *crossover*.

prófago "Cromossomo" de fago inserido como parte da estrutura linear do DNA cromossômico de uma bactéria.

projeto de fármaco baseado na estrutura Utilização de informação básica a respeito de processos celulares e do maquinário para desenvolver fármacos.

projeto genoma Esforço em grande escala, com frequência multilaboratorial, necessário para sequenciar um genoma complexo.

promotora Região reguladora que está a uma curta distância da extremidade 5′ de um gene e que atua como o sítio de ligação para a RNA polimerase.

propriedade Aspecto característico de um organismo, tal como o tamanho, a coloração, a forma ou a atividade enzimática.

proteína acessória Proteína associada à DNA polimerase III de *E. coli*, que não é parte do centro catalítico.

proteína ativadora de catabólito (CAP) Proteína que se une ao cAMP em baixas concentrações de glicose e que se liga ao promotor *lac* para facilitar a ação da RNA polimerase.

proteína de ligação ao boxe TATA (TBP) Fator de transcrição geral que se liga ao boxe TATA e auxilia na atração de outros fatores de transcrição gerais e da RNA polimerase II para promotores eucarióticos.

proteína de ligação unifilamentar (SSB) Proteína que se liga a DNA unifilamentares e que previne que a dupla-hélice seja formada novamente antes da

replicação.

proteína fibrosa Proteína com forma linear, tal como os componentes dos cabelos e dos músculos.

proteína globular Proteína com uma estrutura compacta, tal como uma enzima ou um anticorpo.

proteoma Conjunto completo de genes codificadores de proteínas em um genoma.

proto-oncogene Correspondente celular normal de um gene que pode sofrer mutação e se tornar um oncogene dominante.

prototrófica Linhagem de organismos que proliferará em meio mínimo (*comparar* com **auxotrófica**).

provírus DNA genômico de um retrovírus inserido cromossomicamente.

pseudodominância Súbito aparecimento de um fenótipo recessivo em um heredograma, em virtude da deleção de um gene dominante que o mascarava.

pseudogene Gene inativo por mutação, em relação ao qual não existe correspondente funcional em populações do tipo selvagem.

pseudogene processado Pseudogene que surgiu por meio da transcrição reversa de um mRNA maduro e de sua integração no genoma.

pseudoligação Aparecimento da ligação de dois genes em cromossomos translocados.

purina Tipo de base nitrogenada; as bases purinas no DNA são a adenina e a guanina.

QTL *Ver* ***locus*** **do traço quantitativo**.

quebra bifilamentar Quebra do DNA que cliva o arcabouço açúcar-fosfato de ambos os filamentos da dupla-hélice de DNA.

reação da cadeia de polimerase (PCR) Método *in vitro* para a amplificação de

um segmento de DNA específico, o qual utiliza dois *primers* que hibridizam com extremidades opostas do segmento em polaridade oposta e que, durante ciclos sucessivos, iniciam a replicação exponencial apenas daquele segmento.

rearranjo Produção de cromossomos anormais por meio da quebra e da reunião incorreta de segmentos cromossômicos; exemplos são inversões, deleções e translocações.

rearranjo balanceado Alteração na ordem gênica cromossômica que não remove ou duplica qualquer DNA. As duas classes de rearranjos balanceados são as *inversões* e as *translocações* recíprocas.

rearranjo desbalanceado Rearranjo no qual ocorre um ganho ou uma perda de material cromossômico em um conjunto cromossômico.

receptora Célula bacteriana que recebe o DNA em uma transferência unilateral entre células; exemplos são F^- em uma conjugação ou a célula transduzida em uma transdução mediada por fago.

recombinação (1) Em geral, qualquer processo em uma célula diploide ou parcialmente diploide que gere novas combinações gênicas ou cromossômicas, não observadas anteriormente naquela célula ou em seus genitores. (2) Na meiose, o processo que gera um produto haploide da meiose, cujo genótipo é diferente de qualquer um dos dois genótipos haploides que constituíram o diploide meiótico.

recombinação de fago Produção de genótipos de fagos recombinantes como resultado da infecção dupla de uma célula bacteriana por diferentes genótipos de fagos “parentais”.

recombinação homóloga não alélica (NAHR) *Crossing over* entre unidades homólogas curtas, observado em diferentes *loci* cromossômicos.

recombinação meiótica Recombinação em virtude da distribuição ou do *crossing over* na meiose.

recombinante Refere-se a um organismo ou uma célula individual que apresenta

um genótipo produzido por recombinação.

recomposição alternativa Processo por meio do qual diferentes RNA mensageiros são produzidos a partir do mesmo transcrito primário, por meio de variações no padrão de recomposição do transcrito. Múltiplas "isoformas" de mRNA podem ser produzidas em uma única célula, ou diferentes isoformas podem demonstrar diferentes padrões de expressão tecido-específicos. Se os éxons alternativos estiverem localizados em matrizes de leitura abertas das isoformas de mRNA, diferentes proteínas serão produzidas pelos mRNA alternativos.

recomposição do RNA Reação amplamente observada em eucariotos, que remove íntrons e une éxons no RNA.

rede de haplótipos Rede que demonstra relações entre os haplótipos e as posições das mutações que definem os haplótipos nos ramos.

região 3′ não traduzida (3′ UTR) Região do transcrito de RNA na extremidade 3′ *downstream* do sítio de término da tradução.

região 5′ não traduzida de (5′ UTR) Região do transcrito de RNA na extremidade 5′ *upstream* do sítio de início da tradução.

regiões pseudoautossômicas 1 e 2 Pequenas regiões nas extremidades dos cromossomos sexuais X e Y; elas são homólogas e sofrem pareamento e *crossing over* na meiose.

regra da soma A probabilidade de que um ou outro de dois eventos mutuamente exclusivos irá ocorrer é a soma das suas probabilidades individuais.

regra do produto A probabilidade de dois eventos independentes ocorrerem simultaneamente é o produto das probabilidades individuais.

regra GU-AG Assim denominada em virtude dos dinucleotídios GU e AG estarem quase sempre nas extremidades 5′ e 3′ de íntrons, respectivamente, nas quais são reconhecidos por componentes do spliceossomo.

regulon Genes transcritos de um modo coordenado pela mesma proteína

reguladora (p. ex., fator sigma).

relógio molecular Taxa de substituição constante de aminoácidos em proteínas, ou de nucleotídios em ácidos nucleicos, durante o longo tempo evolutivo.

remoção e união (*splicing*) Reação que remove íntrons e une éxons no RNA.

remodelagem da cromatina Alterações na posição do nucleossomo ao longo do DNA.

reparo genômico global (GGR) Tipo de reparo por excisão de nucleotídio que ocorre em sequências não transcritas.

reparo por excisão de base Uma das diversas vias de reparo por excisão. Nessa via, distorções de pares de bases sutis são reparadas por meio da criação de sítios apurínicos, seguida pela síntese de reparo.

reparo por excisão de nucleotídio acoplado à transcrição (TC-NER) Tipo de reparo por excisão de nucleotídio que é ativado por complexos de transcrição parados e que corrige o dano ao DNA em regiões transcritas do genoma.

repetição de trinucleotídios *Ver* **expansão de trincas**.

replicação conservativa Modelo refutado de síntese de DNA, que sugere que metade das moléculas-filhas de DNA devem apresentar ambos os filamentos compostos por nucleotídios recém-polimerizados.

replicação dispersiva Modelo refutado de síntese de DNA que sugere um entremeado mais ou menos aleatório de segmentos parentais e novos em moléculas-filhas de DNA.

replicação do DNA Processo de síntese de duas cópias idênticas de uma molécula de DNA a partir de uma cópia original.

replicação por círculo rolante Tipo de replicação utilizado por algumas moléculas de DNA circular em bactérias (tais como plasmídios), no qual o círculo aparenta rodar na medida em que desenrola um filamento contínuo.

replicação semiconservativa Modelo estabelecido de replicação do DNA no

qual cada molécula bifilamentar é composta por um filamento parental e um filamento recém-polimerizado.

replissomo Máquina molecular na forquilha de replicação que coordena as diversas reações necessárias para a replicação rápida e precisa do DNA.

repressão de catabólito Inativação de um óperon causada pela presença de grandes quantidades do produto metabólico final do óperon.

repressor Proteína que se liga a um elemento de ação cis, tal como um operador ou um silenciador, prevenindo, assim, a transcrição a partir de um promotor adjacente.

resgate de mutante *Ver* **complementação funcional**.

resposta à seleção (*R*) Quantidade de alteração no valor médio de alguma característica fenotípica entre a geração parental e a geração da progênie, como um resultado da seleção dos genitores.

retrotransposição Mecanismo de transposição caracterizado pelo fluxo reverso de informação do RNA para o DNA.

retrotranspóson Elemento de transposição que utiliza a transcriptase reversa para a transposição por meio de um RNA intermediário. *Ver* **elemento de transposição da classe 1**.

retrotranspóson LTR Tipo de elemento de transposição da classe 1 que termina em repetições terminais longas e codifica diversas proteínas, incluindo a transcriptase reversa.

retrovírus Vírus de RNA que se replica primeiro sendo convertido em um DNA bifilamentar.

revertente Alelo com função do tipo selvagem que surge por meio da mutação de um alelo mutante; causado por uma reversão completa do evento original, ou por uma mutação compensatória em um segundo sítio.

RF *Ver* **fator de liberação**.

RFLP *Ver* **polimorfismo do comprimento do fragmento de restrição**.

ribose Açúcar pentose do RNA.

ribossomo Organela complexa que catalisa a tradução do RNA mensageiro em uma sequência de aminoácidos; composto por proteínas mais rRNA.

ribozima RNA com atividade enzimática – por exemplo, as moléculas de RNA autorrecompostas em *Tetrahymena*.

RISC (complexo de silenciamento induzido por RNA) Complexo proteico de multissubunidades que se associa com os siRNA e que é orientado até um mRNA-alvo por meio da complementaridade de bases. O mRNA-alvo é clivado pela atividade do RISC.

RNA (ácido ribonucleico) Ácido nucleico unifilamentar semelhante ao DNA, mas que apresenta o açúcar ribose em vez do açúcar desoxirribose e uracila em vez de timina como uma das bases.

RNA bifilamentar (dsRNA) Molécula de RNA composta por dois filamentos complementares.

RNA blotting *Ver* ***Northern blot***.

RNA de interação *piwi* (piRNA) RNA transcrito a partir de agrupamentos de pi, que auxilia na proteção da integridade dos genomas de plantas e animais e na prevenção da difusão de elementos de transposição para outros *loci* cromossômicos. Os piRNA restringem os elementos de transposição em animais.

RNA de interferência (RNAi) Sistema em eucariotos para o controle da expressão gênica por meio da ação de siRNA e miRNA. *Ver* **silenciamento gênico**.

RNA funcional Tipo de RNA que desempenha um papel sem ser traduzido.

RNA mensageiro *Ver* **mRNA**.

RNA mundial Denominação de uma teoria popular, de que o RNA deve ter sido o material genético nas primeiras células, tendo em vista que sabidamente apenas o

RNA codifica informações genéticas e catalisa reações biológicas.

RNA não codificador de proteína (ncRNA) RNA que não é traduzido em proteína.

RNA polimerase Enzima que catalisa a síntese de um filamento de RNA a partir de um molde de DNA. Eucariotos apresentam diversas classes de RNA polimerase; os genes estruturais que codificam proteínas são transcritos pela RNA polimerase II.

RNA ribossômico *Ver* **rRNA**.

RNA transportador *Ver* **tRNA**.

RNAi *Ver* **RNA de interferência**.

rRNA (RNA ribossômico) Classe de moléculas de RNA, codificadas no organizador nucleolar, que apresentam um papel integrante (porém pouco compreendido) na estrutura e na função ribossômica.

s *Ver* **coeficiente de seleção**.

S *Ver* **sítios segregantes** ou **diferencial de seleção**.

SDSA *Ver* **pareamento de filamento dependente de síntese**.

segregação adjacente 1 Em uma translocação recíproca, a passagem de um cromossomo translocado e de um normal para cada um dos polos.

segregação alternada Em uma translocação recíproca, a passagem de ambos os cromossomos normais para um polo e de ambos os cromossomos translocados para o outro polo.

segregação citoplasmática Segregação na qual células-filhas geneticamente diferentes têm origem a partir de um genitor que é um *citohet*.

segregação independente *Ver* **segunda lei de Mendel**.

segunda geração filial (F_2) Progênie de um cruzamento entre dois indivíduos da geração F_1.

segunda lei de Mendel (lei da segregação independente) Pares de genes não ligados ou ligados de modo distante, ao segregarem, distribuem-se independentemente na meiose.

seleção (1) Procedimento experimental no qual apenas um tipo específico de mutante consegue sobreviver. (2) Produção de números médios diferentes de progênie, por parte de genótipos diferentes em uma população, como um resultado das diferentes propriedades fenotípicas daqueles genótipos.

seleção artificial Cruzamento de sucessivas gerações por meio da seleção humana deliberada de determinados fenótipos ou genótipos como os genitores de cada geração.

seleção balanceada Seleção natural que resulta em um equilíbrio com frequências alélicas intermediárias.

seleção cumulativa Situação em que a seleção natural promove múltiplas substituições que alteram a função de uma proteína ou de um elemento regulador por meio de rodadas repetidas de mutação e seleção.

seleção direcional Seleção que altera a frequência de um alelo em um sentido constante, seja a favor ou contra a fixação desse alelo.

seleção natural Taxa de reprodução diferencial de diversos tipos em uma população como resultado de distintas características fisiológicas, anatômicas ou comportamentais dos tipos.

seleção negativa Eliminação de um traço deletério de uma população por meio da seleção natural.

seleção positiva Processo por meio do qual um alelo favorável é trazido até uma frequência mais alta em uma população, em virtude de os indivíduos que carreiam aquele alelo apresentarem descendência mais viável do que outros indivíduos.

seleção purificadora Seleção natural que remove variantes deletérias de uma sequência de DNA ou proteína, reduzindo, assim, a diversidade genética.

semiesterilidade (meia-esterilidade) Fenótipo de um organismo heterozigoto em

relação a determinados tipos de aberração cromossômica; expressa-se como um número reduzido de gametas viáveis e, portanto, fertilidade reduzida.

sequência consenso Sequência de nucleotídios de um segmento de DNA que é derivada do alinhamento de sequências semelhantes (seja do mesmo organismo, seja de organismos diferentes) e que determina o nucleotídio mais comum em cada posição.

sequência consenso Sequência de nucleotídios de um segmento de DNA que está de acordo com a maior parte das leituras de sequência do mesmo segmento de indivíduos diferentes.

sequência *contig* Grupo de segmentos clonados sobrepostos.

sequência de ativação *upstream* (UAS) Sequência de DNA de levedura localizada a 5′ do promotor gênico; um fator de transcrição se liga à UAS para regular positivamente a expressão gênica.

sequência de localização nuclear (NLS) Parte de uma proteína necessária para o seu transporte do citoplasma para o núcleo.

sequência de repetição invertida (IR) Sequência observada em uma forma idêntica (porém invertida) – por exemplo, nas extremidades opostas de um transpóson de DNA.

sequência interveniente Íntron; segmento de função amplamente desconhecida dentro de um gene. Esse segmento inicialmente é transcrito, mas o transcrito não é observado no mRNA funcional.

sequência IR *Ver* **sequência de repetição invertida**.

sequência líder Sequência na extremidade 5′ de um mRNA que não é traduzida em proteína.

sequência Shine-Dalgarno Sequência curta no RNA bacteriano que precede o códon AUG de iniciação e que atua para posicionar corretamente esse códon no sítio P do ribossomo pelo pareamento (por meio da complementaridade de bases) com a extremidade 3′ do RNA 16S na subunidade ribossômica 30S.

sequência sinalizadora Sequência aminoterminal de uma proteína secretada; é necessária para o transporte da proteína através da membrana celular.

sequenciamento de RNA Método utilizado para determinar as regiões transcritas de um genoma em alguma população celular específica, amostra tecidual, ou organismo.

sequenciamento de Sanger *Ver* **sequenciamento didesóxi**.

sequenciamento didesóxi (Sanger) Método mais popular de sequenciamento do DNA. Utilizam-se didesoxinucleotídios trifosfatos misturados com nucleotídios trifosfatos padrão para produzir uma série de filamentos de DNA, cuja síntese é bloqueada em diferentes comprimentos. Esse método tem sido incorporado em máquinas de síntese de DNA automatizadas. Também denominado sequenciamento de Sanger em homenagem ao seu inventor, Frederick Sanger.

série alélica Conjunto de alelos conhecidos de um gene. *Ver também* **alelos múltiplos**.

sexo heterogamético Sexo que apresenta cromossomos sexuais heteromórficos (p. ex., XY) e que, portanto, produz dois tipos de gametas diferentes em relação aos cromossomos sexuais.

sexo homogamético Sexo com cromossomos sexuais homólogos (p. ex., XX).

silenciamento epigenético Repressão da expressão de um gene em virtude de sua posição no cromossomo, em vez de por meio de uma mutação na sua sequência de DNA. O silenciamento epigenético pode ser herdado de uma geração celular para a próxima.

silenciamento gênico Gene que não é expresso em virtude de regulação epigenética. Contrariamente aos genes mutantes em virtude de alterações na sequência do DNA, os genes inativados por silenciamento podem ser reativados.

silenciamento gênico pós-transcrição Ocorre quando o mRNA de um gene em particular é destruído ou sua tradução é bloqueada. O mecanismo de silenciamento normalmente envolve RNAi ou miRNA.

silenciamento gênico transcricional Ocorre quando um gene não pode ser transcrito em virtude de estar localizado na heterocromatina.

síndrome de Cockayne Distúrbio genético causado por defeitos no sistema de reparo por excisão de nucleotídio e que leva a sintomas de envelhecimento precoce. Indivíduos com síndrome de Cockayne apresentam uma mutação em uma de duas proteínas que se acredita reconhecerem os complexos de transcrição bloqueados em virtude do dano ao DNA.

síndrome de Down Fenótipo humano anormal, que inclui o retardo mental, em virtude de uma trissomia do cromossomo 21; mais comum em bebês nascidos de mães mais velhas.

síndrome de Klinefelter Fenótipo masculino humano anormal, em virtude de um cromossomo X extra (XXY).

síndrome de Turner Fenótipo feminino humano anormal produzido pela presença de apenas um cromossomo X (XO).

SINE *Ver* **elemento curto intercalado**.

sintenia Situação na qual os genes estão dispostos em blocos semelhantes em diferentes espécies.

síntese de DNA translesão Mecanismo de tolerância a dano em eucariotos, que utiliza *polimerases de bypass* para replicar o DNA além de um sítio de dano.

siRNA *Ver* **pequeno RNA de interferência**.

sistema de reparo de malpareamento Sistema para o reparo de dano ao DNA que já foi replicado.

sistema de reparo por excisão de nucleotídio (NER) Via de reparo por excisão, que quebra as ligações fosfodiéster em cada lado de uma base danificada, removendo aquela base e outras em cada um dos lados, seguida pela replicação de reparo.

sistema seletivo Técnica de seleção mutacional que enriquece a frequência de genótipos específicos (normalmente raros) por meio do estabelecimento de

condições ambientais que previnem o crescimento ou a sobrevivência de outros genótipos.

sistema SOS (de reparo) Processo propenso a erro, por meio do qual uma polimerase de *bypass* realiza a replicação após o dano ao DNA em uma forquilha de replicação bloqueada por meio da inserção de bases inespecíficas.

sítio A *Ver* **sítio de ligação de aminoacil-tRNA**.

sítio alostérico Sítio em uma proteína ao qual uma pequena molécula se liga, causando uma alteração na conformação da proteína, que modifica a atividade de seu sítio ativo.

sítio apurínico Sítio do DNA que perdeu um resíduo purina.

sítio ativo Parte de uma proteína que deve ser mantida em uma forma específica para que a proteína seja funcional – por exemplo, em uma enzima, a parte à qual o substrato se liga.

sítio de inserção λ **(lambda)** Sítio no qual o prófago λ se insere no cromossomo de *E. coli*.

sítio de ligação Região na qual o prófago se integra.

sítio de ligação de aminoacil-tRNA Um sítio no ribossomo que se liga ao aminoacil-tRNA de entrada. O anticódon de cada aminoacil-tRNA de entrada é complementar ao códon do mRNA. Também denominado sítio A.

sítio de peptidil (P) Sítio no ribossomo ao qual um tRNA com a cadeia polipeptídica crescente está ligado.

sítio de saída (E) Sítio no ribossomo no qual o tRNA desacetilado pode ser observado.

sítio E *Ver* **sítio de saída**.

sítio P *Ver* **sítio de peptidil**.

sítios segregantes (*S*) Número de sítios de nucleotídios variáveis ou polimórficos em um conjunto de sequências de DNA homólogas.

SNP *Ver* **polimorfismo de nucleotídio único**.

SNP comum Polimorfismo de nucleotídio único (SNP) em relação ao qual o alelo menos comum ocorre a uma frequência de aproximadamente 5% ou mais.

SNP raro Polimorfismo de nucleotídio único (SNP) em relação ao qual o alelo menos comum ocorre a uma frequência inferior a 5%.

snRNA *Ver* **pequeno RNA nuclear**.

sonda Segmento de ácido nucleico marcado, que pode ser utilizado para identificar moléculas de DNA específicas que apresentam a sequência complementar, normalmente por meio de autorradiografia ou fluorescência.

Southern blot Transferência de fragmentos de DNA separados por eletroforese, de um gel para uma folha absorvente, tal como um papel; essa folha em seguida é imersa em uma solução que contém uma sonda marcada, que se ligará a um fragmento de interesse.

spliceossomo Complexo de processamento ribonucleoproteico que remove íntrons de mRNA eucarióticos.

SSB *Ver* **proteína de ligação unifilamentar**.

SSLP *Ver* **polimorfismo de comprimento de sequência simples**.

subfuncionalização Via de duplicação e mutação gênica que produz *parálogos* com funções complementares.

substituição gênica Inserção de um transgene geneticamente modificado no lugar de um gene residente; com frequência realizada por meio de um *crossover* duplo.

substituição não conservativa Substituição de um par de nucleotídios em uma região codificadora de proteína, que leva à substituição de um aminoácido por outro que apresenta propriedades químicas diferentes.

substituição não sinônima Alteração na sequência de DNA codificadora de proteína que causa a alteração de um aminoácido.

substituição sinônima *Ver* **mutação sinônima**.

subunidade Conforme utilizada no Capítulo 9, um polipeptídio único em uma proteína que contém múltiplos polipeptídios.

sulco menor Menor dos dois sulcos na dupla-hélice de DNA.

sulco principal Maior dos dois sulcos na dupla-hélice do DNA.

supercontig *Ver* **arcabouço** (2).

supressor Mutação secundária que pode cancelar o efeito de uma mutação primária, resultando em fenótipo do tipo selvagem.

tautomerização *Ver* **mudança tautomérica**.

taxa de mutação (μ) Probabilidade de que uma cópia de um alelo seja alterada para algum outro tipo alélico em uma geração.

TBP *Ver* **proteína de ligação ao boxe TATA**.

TC-NER *Ver* **reparo por excisão de nucleotídio acoplado à transcrição**.

tecnologia do DNA Técnicas coletivas para a obtenção, a amplificação e a manipulação de fragmentos de DNA específicos.

telomerase Enzima que, com a utilização de um pequeno RNA especial como molde, adiciona unidades repetitivas às extremidades de cromossomos lineares para prevenir o encurtamento após a replicação.

telômero Ponta, ou extremidade, de um cromossomo.

temperatura permissiva Temperatura na qual um alelo mutante sensível à temperatura é expresso do mesmo modo que o alelo do tipo selvagem.

temperatura restritiva Temperatura na qual uma mutação sensível à temperatura expressa o fenótipo mutante.

terapia gênica Correção de uma deficiência genética em uma célula por meio da adição de um novo DNA e sua inserção no genoma. Diferentes técnicas apresentam o potencial de realizar a terapia gênica apenas em tecidos somáticos ou, alternativamente, de corrigir a deficiência genética no zigoto, assim corrigindo também a linhagem germinativa.

terminação Último estágio da transcrição; resulta na liberação do RNA e da RNA polimerase do molde de DNA.

término O fim representado pelo último monômero adicionado na síntese unidirecional de um polímero, tal como RNA ou um polipeptídio.

testador Organismo homozigoto em relação a um ou mais alelos recessivos; utilizado em um cruzamento-teste.

teste de Ames Modo de testar se um composto químico é mutagênico por meio da exposição de linhagens bacterianas mutantes especiais ao produto formado pela digestão daquele composto por parte de um extrato hepático e, em seguida, da contagem do número de colônias. Apenas mutações novas, presumivelmente produzidas pelo composto, podem produzir revertentes para o tipo selvagem capazes de formar colônias.

teste de complementação Teste para determinar se duas mutações estão em genes diferentes (elas se complementam) ou no mesmo gene (elas não se complementam).

teste de flutuação Teste utilizado em microrganismos para estabelecer a natureza aleatória da mutação ou para medir as taxas de mutação.

teste di-híbrido Método para a detecção de interações de proteínas, tipicamente realizado em leveduras.

teste do qui-quadrado (χ^2) Teste estatístico utilizado para determinar a probabilidade de obtenção de proporções observadas ao acaso, sob uma hipótese específica.

tétrade (1) Quatro cromátides homólogas em um feixe na primeira prófase meiótica e na metáfase. (2) Quatro células haploides que são produtos de uma única meiose.

tetraploide Célula que apresenta quatro conjuntos cromossômicos; um organismo composto pelas referidas células.

timina (T) Base pirimidina que pareia com a adenina.

tipo selvagem Genótipo ou fenótipo que é observado na natureza ou no estoque de laboratório padrão em relação a um determinado organismo.

Tn *Ver* **transpóson**.

***toolkit* genética** Conjunto de genes responsável pela regulação do desenvolvimento animal, amplamente composto por membros de vias de sinalização celular e fatores de transcrição.

topoisomerase Enzima que consegue cortar e formar novamente as espinhas dorsais de polinucleotídios no DNA para possibilitar que ele assuma uma configuração mais relaxada.

traço Mais ou menos sinônimo de fenótipo.

traço categórico Traço em relação ao qual os indivíduos podem ser classificados em grupos discretos ou descontínuos, tal como hastes altas *versus* baixas nas ervilhas de Mendel.

traço complexo Traço que exibe herança complexa.

traço contínuo Traço que pode adotar um número possivelmente infinito de estados em uma variação contínua, tal como a altura em seres humanos.

traço de limiar Traço categórico em relação ao qual a expressão dos diferentes estados fenotípicos depende de uma combinação de múltiplos fatores genéticos e/ou ambientais que posicionam um indivíduo acima ou abaixo de um valor crítico em relação à expressão do traço.

traço merístico Traço de contagem, que adota uma diversidade de valores discretos.

traço quantitativo Qualquer traço que exibe herança complexa em virtude de ser controlado por uma mistura de fatores genéticos e/ou ambientais.

tradução Produção de um polipeptídio mediada por ribossomo e tRNA, cuja sequência de aminoácidos é derivada da sequência de códons de uma molécula de mRNA.

transcrição Síntese de RNA a partir de um molde de DNA.

transcriptase reversa Enzima que catalisa a síntese de um filamento de DNA a partir de um molde de RNA.

transcrito Molécula de RNA copiada do filamento-molde de DNA pela RNA polimerase.

transcrito primário (pré-mRNA) RNA eucariótico antes de seu processamento.

transdução Movimento de genes de um doador bacteriano para um receptor bacteriano com um fago como vetor.

transdução especializada Situação na qual um fago em particular transduzirá apenas regiões específicas do cromossomo bacteriano.

transdução generalizada Capacidade de determinados fagos de transduzir qualquer gene no cromossomo bacteriano.

transformação Modificação direcionada de um genoma por meio da aplicação externa do DNA de uma célula de genótipo diferente.

transformação dupla Transformação simultânea por dois marcadores doadores diferentes.

transgene Gene que foi modificado por meio de técnicas de DNA recombinante aplicadas externamente e reintroduzido no genoma por meio de transformação da linhagem germinativa.

transição Tipo de substituição de par de nucleotídios, no qual uma purina substitui outra purina, ou no qual uma pirimidina substitui outra pirimidina – por exemplo, G-C para A-T.

transição alostérica Alteração da conformação de uma proteína em outra.

translocação Realocação de um segmento cromossômico para uma posição diferente no genoma.

transmissão horizontal Herança do DNA de outro membro da mesma geração.

transmissão vertical Herança do DNA de um membro de uma geração anterior.

transpor Mover-se de um local no genoma para outro; diz-se de um elemento genético móvel.

transposase Enzima codificada por elementos de transposição que são submetidos à transposição conservativa.

transposição Processo por meio do qual elementos genéticos móveis se movimentam de um local no genoma para outro.

transposição conservativa Mecanismo de transposição que movimenta um elemento móvel até uma nova localização no genoma na medida em que ele é removido de sua localização anterior.

transposição replicativa Mecanismo de transposição que gera um novo elemento de inserção integrado em algum outro local no genoma, ao mesmo tempo que deixa o elemento original no seu sítio de inserção original.

transpóson (Tn) Parte móvel do DNA que é flanqueada por sequências repetidas terminais e que tipicamente contém genes que codificam funções de transposição. Os transpósons bacterianos podem ser simples ou compostos.

transpóson composto Tipo de elemento de transposição bacteriano que contém uma diversidade de genes que estão localizados entre dois elementos de sequência de inserção (IS) quase idênticos.

transpóson de DNA *Ver* **elemento DNA**.

transpóson simples Tipo de elemento de transposição bacteriano, que contém uma diversidade de genes que estão localizados entre sequências repetidas curtas invertidas.

transversão Tipo de substituição de par de nucleotídios no qual uma pirimidina substitui uma purina, ou vice-versa – por exemplo, G-C para T-A.

triagem Procedimento de mutagênese no qual essencialmente toda a progênie que sofreu mutação é recuperada e avaliada individualmente em relação ao fenótipo mutante; com frequência o fenótipo desejado é marcado de alguma maneira para possibilitar a sua detecção.

trinca Três pares de nucleotídios que compõem um códon.

triploide Célula que apresenta três conjuntos cromossômicos ou um organismo composto por tais células.

trissômico Basicamente um diploide com um cromossomo extra de um tipo, que produz um número de cromossomos do tipo $2n + 1$.

trivalente Refere-se ao arranjo de pareamento meiótico de três homólogos em um triploide ou trissômico.

tRNA (RNA transportador ou de transferência) Classe de pequenas moléculas de RNA que carregam aminoácidos específicos para o ribossomo durante a tradução; um aminoácido é inserido na cadeia polipeptídica crescente quando o anticódon do tRNA correspondente pareia com um códon no mRNA que está sendo traduzido.

tRNA carregado Molécula de RNA transportador com um aminoácido ligado à sua extremidade 3′. Também denominado aminoacil-tRNA.

U *Ver* **uracila; uridina**.

u.m. *Ver* **unidade de mapa**.

UAS *Ver* **sequência de ativação *upstream***.

ubiquitina Proteína que, quando unida como uma cadeia multicópias a outra proteína, marca essa proteína para a degradação por uma protease denominada proteassomo 26S. A adição de resíduos de ubiquitina únicos a uma proteína pode alterar as interações proteína-proteína, como no caso de PCNA e polimerases de *bypass*.

ubiquitinização Processo de adição de cadeias de ubiquitina a uma proteína-alvo para a degradação.

unidade de mapa (u.m.) "Distância" entre dois pares de genes ligados, na qual 1% dos produtos da meiose é recombinante; uma unidade de distância em um mapa de ligação.

unidade de mapa genético (u.m.) Distância no mapa cromossômico, que corresponde à frequência de recombinantes de 1%.

univalente Cromossomo meiótico não pareado único, conforme observado com frequência em trissômicos e triploides.

upstream Refere-se a uma sequência de DNA ou RNA localizada no lado 5′ de um ponto de referência.

uracila (U) Base pirimidina no RNA no lugar da timina observada no DNA.

uridina (U) Nucleosídio que apresenta uracila como sua base.

UTR *Ver* **região 3′ não traduzida; região 5′ não traduzida**.

valor adaptativo absoluto Número de progênie que um indivíduo apresenta.

valor adaptativo relativo Medida da adaptabilidade de um indivíduo ou um genótipo em relação a algum outro indivíduo ou genótipo, normalmente o indivíduo ou genótipo mais adaptado na população.

Valor *C* Conteúdo de DNA de um genoma haploide.

valor produtivo Parte do desvio de um indivíduo da média da população que ocorre em virtude de efeitos aditivos e que é transmitida para a sua progênie.

variação genética aditiva Parte da variância genética em relação a um traço em uma população que é transmitida de modo previsível dos genitores para a progênie.

variação no número de cópias (CNV) Variação de um grande segmento de DNA entre cromossomos homólogos, causada por diferenças no número de cópias em *tandem* de um gene único ou de múltiplos genes.

variância Medida estatística utilizada para mensurar o quanto os valores de traço dos indivíduos se desviam da média da população.

variância ambiental Parte da variação fenotípica entre os indivíduos em uma população que ocorre em virtude dos diferentes ambientes que os indivíduos

vivenciaram.

variância genética Parte da variação fenotípica entre indivíduos em uma população que ocorre em virtude das diferenças genéticas entre os indivíduos.

variegação por efeito de posição (PEV) Variegação causada pela inativação de um gene em algumas células por meio de sua justaposição anormal com a heterocromatina.

vetor *Ver* **vetor de clonagem**.

vetor de clonagem Na clonagem, o plasmídio ou cromossomo de fago utilizado para carrear o segmento de DNA clonado.

vigilância do genoma Coleção de mecanismos que reconhece e destrói ácidos nucleicos invasores ou transpósons ativos. *Ver também* **crRNA** e **piRNA**.

vigor híbrido Situação na qual a F_1 é maior ou mais saudável do que as suas duas linhagens parentais puras diferentes.

vírus Partícula composta por ácido nucleico e proteína, que precisa infectar uma célula viva para replicar e se reproduzir.

VNTR *Ver* **número variável de repetições em *tandem***.

XP *Ver* **xeroderma pigmentoso**.

xeroderma pigmentoso (XP) Distúrbio causado por mutações no sistema de reparo por excisão de nucleotídio acoplado à transcrição, que leva ao frequente desenvolvimento de cânceres de pele.

zigoto Célula formada pela fusão de um ovócito e um espermatozoide; a célula diploide única que se dividirá mitoticamente para criar um organismo diploide diferenciado.

Respostas dos Problemas Selecionados

Esta seção inclui respostas selecionadas dos Problemas Básicos e Problemas Desafiadores de todos os capítulos, com exceção do Capítulo 1. As respostas dos problemas do Capítulo 1 não estão incluídas aqui pois são questões de discussão.

Capítulo 2

15. A PFGE separa as moléculas de DNA pelo tamanho. Quando o DNA é cuidadosamente isolado de *Neurospora* (que apresenta sete cromossomos diferentes), devem ser produzidas sete bandas com a utilização dessa técnica. De modo semelhante, a ervilha apresenta sete cromossomos diferentes e produzirá sete bandas (os cromossomos homólogos migrarão em conjunto como uma banda única).

18. A função-chave da mitose é gerar duas células-filhas geneticamente idênticas à célula parental original.

22. Na medida em que as células se dividem por mitose, cada cromossomo é composto por cromátides-irmãs idênticas, que são separadas para formar células-filhas geneticamente idênticas. Embora a segunda divisão da meiose aparente ser um processo semelhante, as cromátides-"irmãs" provavelmente serão diferentes uma da outra. A recombinação nos estágios meióticos mais iniciais apresentará regiões de DNA permutadas entre os cromossomos-irmãos e não irmãos, de tal modo que as duas células-filhas dessa divisão tipicamente não são geneticamente idênticas.

26. Sim. Metade de nossa constituição é derivada de cada genitor, metade da constituição genética de cada genitor é derivada de metade de cada um de seus genitores etc.

30. (d) Sinapse (pareamento cromossômico).

35. A proporção da progênie é de aproximadamente 3:1, indicando cruzamento clássico entre heterozigotos. Tendo em vista que a coloração preta (*B*) é dominante em relação à branca (*b*):

Genitores:	*B/b* × *B/b*
Progênie:	3 pretas:1 branca (1 *B/B*:2 *B/b*:1 *b/b*)

39. O fato de que aproximadamente metade da progênie F_1 é de mutantes sugere que a mutação que resulta em três cotilédones é dominante e que o mutante original era heterozigoto. Se *C* = Alelo mutante e *c* = Alelo do tipo selvagem, o

cruzamento é como segue:

P	*C/c* × *c/c*
F_1	*C/c*, três cotilédones
	c/c, dois cotilédones

44.*p* (o filho apresenta galactosemia) = *p* (John é *G/g*) × *p* (Martha é *G/g*) × *p* (ambos os genitores transmitiram *g* para o filho) = (2/3) (1/4) (1/4) = 2/48 = 1/24.

50.a. O distúrbio aparenta ser dominante, tendo em vista que todos os indivíduos afetados apresentam um genitor afetado. Se o traço fosse recessivo, então I-1, II-2, III-1 e III-8 deveriam, todos, ser portadores (heterozigotos em relação ao alelo raro).

b. Com a presunção de dominância, os genótipos são:

I:	*d/d, D/d*
II:	*D/d, d/d, D/d, d/d*
III:	*d/d, D/d, d/d, D/d, d/d, d/d, D/d, d/d*
IV:	*D/d, d/d, D/d, d/d, d/d, d/d, d/d, D/d, d/d*

c. A probabilidade de um filho afetado (*D/d*) é igual a 1/2 e a probabilidade de um não afetado (*d/d*) é igual a 1/2. Portanto, a chance de ter quatro filhos não afetados (tendo em vista que cada filho é um evento independente) é de (1/2) × (1/2) × (1/2) × (1/2) = 1/16.

56. a. Os filhos do sexo masculino herdam o cromossomo X de suas mães. A mãe apresenta os lóbulos das orelhas livres; o filho apresenta os lóbulos das orelhas presos. Se o alelo em relação aos lóbulos das orelhas livres for dominante e o alelo em relação aos presos for recessivo, então a mãe poderá ser heterozigota em relação a esse traço e o gene pode ser ligado ao X.

b. Não é possível, a partir dos dados fornecidos, decidir qual alelo é dominante. Se os lóbulos das orelhas presos representarem traço dominante, então o pai seria heterozigoto e o filho apresentaria uma chance de 50% de herdar o alelo dominante dos lóbulos presos. Se os lóbulos das orelhas presos forem traço recessivo, então o traço poderia ser autossômico ou ligado ao X, mas, em qualquer caso, a mãe seria heterozigota.

60. Estabeleça que *H* = Hipofosfatemia e *h* = Normal. O cruzamento é *H/Y* × *h/h*, produzindo *H/h* (mulheres) e *h/Y* (homens). A resposta é 0%.

65. a. X^C/X^c, X^c/X^c

b. *p* (daltônica) × *p* (homem) = (1/2) (1/2) = 1/4

c. As meninas serão 1 normal (X^C/X^c):1 daltônica (X^c/X^c).

d. O cruzamento é $X^C/X^c \times X^c/Y$, produzindo 1 normal:1 daltônico para ambos os sexos.

73. a. O heredograma sugere que o alelo que causa cabelos ruivos é recessivo, tendo em vista que a maior parte dos indivíduos ruivos descendem de genitores sem esse traço.

b. A observação das pessoas ao nosso redor faz com que o alelo pareça um pouco raro.

77. Observe que apenas os homens são afetados e que, em todos os casos, com exceção de um, o traço pode ser acompanhado pelo lado feminino. Entretanto, existe o exemplo de um homem afetado, que apresenta filhos afetados. Se o traço for ligado ao X, a esposa desse homem deve ser portadora. Dependendo de quão raro esse traço é na população geral, isso pode ser improvável, sugerindo que o distúrbio é causado por um alelo autossômico dominante com expressão limitada aos homens.

Capítulo 3

13. O genótipo das células-filhas será idêntico àquele da célula original: (f) *A/a*; *B/b*.

18. A mitose produz células-filhas que apresentam o mesmo genótipo da célula original: *A/a*; *B/b*; *C/c*.

21. Seus filhos precisarão herdar o cromossomo 4 que contém o satélite (probabilidade = 1/2), o cromossomo 7 de coloração anormal (probabilidade = 1/2) e o cromossomo Y (probabilidade = 1/2). Para herdar todos os três, a probabilidade é de (1/2) (1/2) (1/2) = 1/8.

26. Com a presunção de segregação independente e relações de dominância e recessividade simples de todos os genes, o número de classes genotípicas esperadas a partir da autofecundação de uma planta heterozigota em relação a *n* pares de genes é 3*n* e o número de classes fenotípicas esperadas é 2*n*.

29. a. e **b.** O cruzamento 2 indica que roxo (*G*) é dominante em relação a verde (*g*) e o cruzamento 1 indica que *cut* (*P*) é dominante em relação a *potato* (*p*).

Cruzamento 1:	*G/g*; *P/p* × *g/g*; *P/p*	Existem 3 *cut*:1 *potato* e 1 roxo:1 verde
Cruzamento 2:	*G/g*; *P/g* × *G/g*; *p/p*	Existem 3 roxos:1 verde e 1 *cut*:1 *potato*
Cruzamento 3:	*G/G*; *P/p* × *g/g*; *P/p*	Não existem verdes e existem 3 *cut*:1 *potato*
Cruzamento 4:	*G/g*; *P/P* × *g/g*; *p/p*	Não existem *potato* e existe 1 roxo:1 verde
Cruzamento 5:	*G/g*; *p/p* × *g/g*; *P/p*	Existe 1 *cut*:1 *potato* e existe 1 roxo:1 verde

34. Os cruzamentos são:

Cruzamento 1:	fêmea intermitente × macho do tipo selvagem → toda a progênie intermitente
Cruzamento 2:	fêmea do tipo selvagem × macho intermitente → toda a progênie do tipo selvagem

O mtDNA é herdado apenas da "fêmea" em *Neurospora*.

40. a. Deve haver nove classes, correspondentes a 0, 1, 2, 3, 4, 5, 6, 7 e 8 "doses".

b. Deve haver 13 classes, correspondentes a 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12 "doses".

49. a. e **b.** Inicie com quaisquer duas das três linhagens e realize o seu cruzamento. Se, por exemplo, você iniciar com *a/a*; *B/B*; *C/C* × *A/A*; *b/b*; *C/C*, toda progênie será *A/a*; *B/b*; *C/C*. O cruzamento de duas delas produzirá:

9	*A/–*; *B/–*; *C/C*
3	*a/a*; *B/–*; *C/C*
3	*A/–*; *b/b*; *C/C*
1	*a/a*; *b/b*; *C/C*

O genótipo *a/a*; *b/b*; *C/C* apresenta dois dos genes em um estado homozigoto recessivo e é observado em 1/16 da descendência. Se tal genótipo fosse cruzado com *A/A*; *B/B*; *c/c*, toda a progênie seria *A/a*; *B/b*; *C/c*. O cruzamento de dois deles (ou a "autofecundação") levaria a uma proporção de 27:9:9:9:3:3:3:1 e 1/64 da progênie seria o desejado *a/a*; *b/b*; *c/c*.

Existem diversos caminhos diferentes para a obtenção de *a/a*; *b/b*; *c/c*, mas o apresentado aqui requer apenas quatro cruzamentos.

56. a. Estabeleça que *B* = Braquidáctilo, *b* = Normal, *T* = Degustador e *t* = Não degustador. Os genótipos do casal são *B/b*; *T/t* em relação ao homem e *b/b*; *T/t* em relação à mulher.

b. Para que todos os quatro filhos sejam braquidáctilos, $p = (1/2)^4 = 1/16$.

c. Para que nenhum dos quatro filhos seja braquidáctilo, $p = (1/2)^4 = 1/16$.

d. Para que todos sejam degustadores, $p = (3/4)^4 = 81/256$.

e. Para que todos sejam não degustadores, $p = (1/4)^4 = 1/256$.

f. Para que todos sejam degustadores braquidáctilos, $p = (1/2 \times 3/4)^4 = 81/4.096$.

g. A probabilidade de não ser um degustador braquidáctilo é de 1 – (a probabilidade de ser um degustador braquidáctilo) ou $1 – (1/2 \times 3/4) = 5/8$. A probabilidade de que todos os quatro filhos sejam degustadores não

braquidáctilos é de $(5/8)^4 = 625/4.096$.

h. A probabilidade de que no mínimo um seja um degustador braquidáctilo é de 1 – (a probabilidade de nenhum ser um degustador braquidáctilo), ou 1 – $(5/8)^4$.

Capítulo 4

13. P *A d/A d* × *a D/a D*

F_1	*A d/a D*	
F_2	1 *A d/A d*	fenótipo: *A d*
	2 *A d/a D*	fenótipo: *A D*
	1 *a D/a D*	fenótipo: *a D*

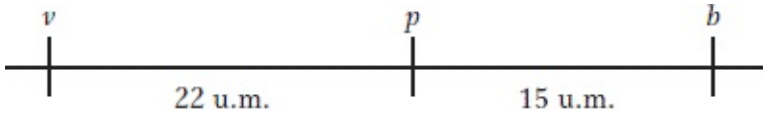
16. Tendo em vista que apenas os tipos parentais são recuperados, os dois genes devem estar fortemente ligados e a recombinação deve ser muito rara. Saber quanto da progênie estamos observando forneceria uma indicação do quão próximos os genes estão.

21. a. Os três genes estão ligados.

b. Uma comparação dos parentais (mais frequentes) com os *crossing overs* duplos (menos frequentes) revela que a ordem dos genes é *v p b*. Existiam 2.200 recombinantes entre *v* e *p* e 1.500 entre *p* e *b*. A fórmula geral para unidades de mapa é:

u.m. = 100% (número de recombinantes)/número total de progênie

Portanto, as unidades de mapa entre *v* e *p* = 100% (2.200)/10.000 = 22 u.m., e as unidades de mapa entre *p* e *b* = 100% (1.500)/10.000 = 15. u.m. O mapa é:

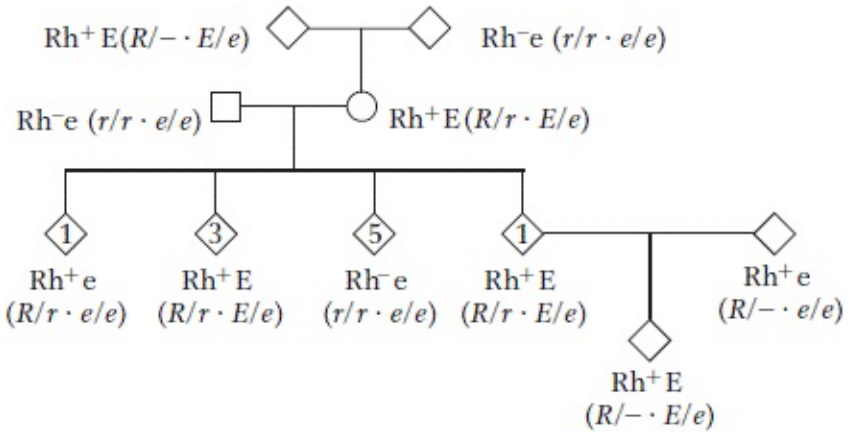


c. I = 1 – *crossing overs* duplos observados/*crossing overs* duplos esperados

= 1 – 132/(0,22) (0,15) (10.000)

= 1 – 0,4 = 0,6

27. a.



b. Sim.

c. Dominante.

d. Conforme desenhado, o heredograma indica uma ligação. Se não estiverem ligados, esperamos que os fenótipos dos 10 filhos devam apresentar uma proporção de 1:1:1:1 de Rh^+ E, Rh^+ e, Rh^- E e Rh^- e. Existem realmente cinco Rh^- e, quatro Rh^+ E e um Rh^+ e. Se estiverem ligados, esse último fenótipo representaria um recombinante e a distância entre os dois genes seria de 100% (1/10) = 10 u.m. Entretanto, não existem dados suficientes para amparar fortemente tal conclusão.

33. a. Se os genes não estiverem ligados, o cruzamento é:

P	*hyg/hyg*; *her/her* × *hyg*$^+$*/hyg*$^+$; *her*$^+$*/her*$^+$
F_1	*hyg*$^+$*/hyg*; *her*$^+$*/her* × *hyg*$^+$*/hyg*; *her*$^+$*/her*
F_2	9/16 *hyg*$^+$*/–*; *her*$^+$*/–*
	3/16 *hyg*$^+$*/–*; *her/her*
	3/16 *hyg/hyg*; *her*$^+$*/–*
	1/16 *hyg/hyg*; *her/her*

Assim, espera-se que apenas 1/16 (ou 6,25%) das sementes germine.

b. e **c.** Não. Mais do que o dobro das sementes esperadas germinou; assim, presume-se que os genes estão ligados. O cruzamento então é:

P	*hyg her/hyg her* × *hyg*$^+$ *her*$^+$*/hyg*$^+$ *her*$^+$
F_1	*hyg*$^+$ *her*$^+$*/hyg her* × *hyg*$^+$ *her*$^+$*/hyg her*
F_2	13% *hyg her/hyg her*

Tendo em vista que essa classe representa a combinação de dois cromossomos parentais, ela é igual a:

$$p\ (hyg\ her) \times p\ (hyg\ her) = (\tfrac{1}{2}\ \text{parentais})^2 = 0{,}13$$

e:

$$\text{parentais} = 0{,}72$$

Assim:

$$\text{recombinantes} = 1 - 0{,}72 = 0{,}28$$

Portanto, um cruzamento-teste de $hyg^+\ hyg^+/hyg\ her$ dará:

36%	$hyg^+\ her^+/hyg\ her$
36%	$hyg\ her/hyg\ her$
14%	$hyg^+\ her/hyg\ her$
14%	$hyg\ her^+/hyg\ her$

e 36% da progênie crescerá (a classe *hyg her/hyg her*).

37. A fórmula para este problema é $f(\text{i}) = e^{-m}m^i/i!$, na qual $m = 2$ e $i = 0$, 1 ou 2.

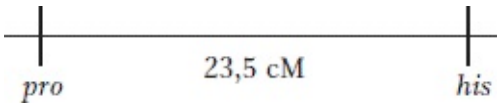
a. $f(0) = e^{-2}2^0/0! = e^{-2} = 0{,}135$, ou 13,5%

b. $f(1) = e^{-2}2^1/1! = e^{-2}(2) = 0{,}27$, ou 27%

c. $f(2) = e^{-2}2^2/2! = e^{-2}(2) = 0{,}27$, ou 27%

43. a. O cruzamento foi *pro* + × + *his*, que torna a primeira classe de tétrades DNP (6 ditipos não parentais), a segunda classe de tétrades T (82 tetratipos) e a terceira classe de tétrades DP (112 ditipos parentais). Quando DP >> DNP, você sabe que os dois genes estão ligados.

b. A distância de mapa pode ser calculada utilizando-se a fórmula FR = [DNP + (1/2)T] 100%. Nesse caso, a frequência de DNP é 6/200, ou 3%, e a frequência de T é 82/200, ou 41%. A distância de mapa entre esses dois *loci*, portanto, é de 23,5 cM.



c. Para corrigir em relação a múltiplos *crossing overs*, pode ser utilizada a fórmula de Perkins. Portanto, a distância de mapa = (T + 6 DNP) 50% ou (0,41 + 0,18) 50% = 29,5 cM.

47. a. O cruzamento é *WeF/WeF* × *wEf/wEf* e a F_1 é *WeF/wEf*. A progênie *ww ee ff* originada a partir de um cruzamento-teste dessa F_1 deve ter herdado um dos cromossomos recombinantes de *crossover* duplo (*w e f*). Com a presunção de ausência de interferência, a porcentagem esperada de *crossing overs* duplos é de 8% × 24% = 1,92%, metade da qual é 0,96%.

b. A obtenção de uma progênie *ww ee ff* a partir de uma autofecundação dessa F_1 requer a herança independente de dois cromossomos *w e f* duplamente recombinantes. As suas chances de ocorrência, com base na resposta da parte *a* deste problema, são de 0,96 × 0,96 = 0,009%.

53. A resposta curta é que os resultados nos informam pouco a respeito da ligação. Embora o número de recombinantes (3) seja inferior ao número de parentais (5), não se pode ter confiança no fato de que a FR é < 50%. O principal problema é que o tamanho da amostra é pequeno e, assim, apenas um indivíduo a mais ou a menos em uma classe genotípica pode afetar dramaticamente as proporções. Até mesmo o teste de qui-quadrado não é confiável nos referidos pequenos tamanhos de amostra. Provavelmente *é* seguro dizer que não existe uma ligação forte, tendo em vista que diversos recombinantes foram observados em uma amostra relativamente pequena. Entretanto, não se pode distinguir entre a ligação mais distante e a distribuição independente. É necessário um tamanho de amostra maior.

58. Os dados fornecidos em relação a cada um dos cruzamentos-teste de três pontos podem ser utilizados para determinar a ordem dos genes quando se percebe que as classes recombinantes mais raras são o resultado de eventos de *crossing overs* duplo. Uma comparação desses cromossomos com os tipos "parentais" revela que os alelos que foram permutados representam o gene no meio.

Por exemplo, no conjunto de dados 1, os fenótipos mais comuns (+ + + e *a b c*) representam as combinações de alelos parentais. Uma comparação desses fenótipos com os fenótipos mais raros desse conjunto de dados (+ *b c* e *a* + +) indica que o gene *a* é recombinante e que deve estar no meio. A ordem dos genes é *b a c*.

Em relação ao conjunto de dados 2, + *b c* e *a* + + (os parentais) devem ser

comparados com + + + e *a b c* (os recombinantes mais raros) para indicar que o gene *a* está no meio. A ordem dos genes é *b a c*.

Em relação ao conjunto de dados 3, compare + *b* + e *a* + *c* com *a b* + e + + *c*, que fornece a ordem de genes *b a c*.

Em relação ao conjunto de dados 4, compare + + *c* e *a b* + com + + + e *a b c*, que fornece a ordem de genes *a c b*.

Em relação ao conjunto de dados 5, compare + + + e *a b c* com + + *c* e *a b* +, que fornece a ordem de genes *a c b*.

64. a. O cruzamento 1 é reduzido para:

P	*A/A · B/B · D/D* × *a/a · b/b · d/d*
F_1	*A/a · B/b · D/d* × *a/a · b/b · d/d*

A progênie do cruzamento-teste indica que esses três genes estão ligados (CO = *Crossing over*; DCO = *Crossing over* duplo).

Progênie do cruzamento-teste	*A B D*	316	parental
	a b d	314	parental
	A B d	31	CO *B-D*
	a b D	39	CO *B-D*
	A b d	130	CO *A-B*
	a B D	140	CO *A-B*
	A b D	17	DCO
	a B d	13	DCO

A-B:	100% (130 + 140 + 17 + 13)/1.000 = 30 u.m.
B-D:	100% (31 + 39 + 17 + 13)/1.000 = 10 u.m.

O cruzamento 2 é reduzido a:

P	*A/A · C/C · E/E* × *a/a · c/c · e/e*
F_1	*A/a · C/c · E/*e × *a/a · c/c · e/e*

A progênie do cruzamento-teste indica que esses três genes estão ligados.

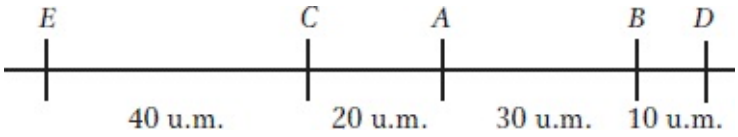
Progênie do cruzamento-teste	*A C E*	243	parental
	a c e	237	parental

A c e	62	CO *A-C*
a C E	58	CO *A-C*
A C e	155	CO *C-E*
a c E	165	CO *C-E*
a C e	46	DCO
A c E	34	DCO

A-B: 100% (62 + 58 + 46 + 34)/1.000 = 20 u.m.

B-D: 100% (155 + 165 + 46 + 34)/1.000 = 40 u.m.

O mapa que acomoda todos os dados é:



b. Interferência (I) = 1 – [(DCO observado)/(DCO esperado)]

Em relação ao cruzamento 1: I = 1 – {30/[(0,30) (0,10) (1.000)]} = 1 – 1 = 0, nenhuma interferência

Em relação ao cruzamento 2: I = 1 – {80/[(0,20) (0,40) (1.000)]} = 1 – 1 = 0, nenhuma interferência

69. a. e **b.** Os dados amparam a segregação independente de dois genes (denomine-os *arg1* e *arg2*). O cruzamento se torna *arg1*; *arg2*$^+$ × *arg1*$^+$; *arg2* e as tétrades resultantes são:

4:0 (DP)	3:1 (T)	2:2 (DNP)
arg1; *arg2*$^+$	*arg1*; *arg2*$^+$	*arg1*; *arg2*
arg1; *arg2*$^+$	*arg1*$^+$; *arg2*	*arg1*; *arg2*
arg1$^+$; *arg2*	*arg1*; *arg2*	*arg1*$^+$; *arg2*$^+$
arg1$^+$; *arg2*	*arg1*$^+$; *arg2*$^+$	*arg1*$^+$; *arg2*$^+$

Tendo em vista que DP = DNP, os genes não estão ligados.

Capítulo 5

19. Uma linhagem Hfr apresenta o fator de fertilidade F integrado no cromossomo. Uma linhagem F^+ apresenta o fator de fertilidade livre no citoplasma. Uma linhagem F^- apresenta ausência do fator de fertilidade.

23. Embora os experimentos de cruzamento interrompido forneçam a ordem dos genes, ela será apenas em relação a marcadores razoavelmente distantes. Portanto, a mutação não pode ser localizada com precisão por meio dessa técnica. A transdução generalizada fornecerá informações a respeito de marcadores muito próximos, o que a torna uma escolha inadequada para os experimentos iniciais, em virtude da quantidade maciça de triagens que precisariam ser realizadas. Em conjunto, as duas técnicas possibilitam, primeiramente, a localização do mutante (cruzamento interrompido) e, em segundo lugar, a determinação precisa da localização do mutante (transdução generalizada) na região geral.

28. A melhor explicação é que o fator F integrado de Hfr forma uma alça fora do cromossomo bacteriano anormalmente e agora é um F′ que contém o gene pro^+. Esse F′ é rapidamente transferido para as células F^-, convertendo-as em pro^+ (e F^+).

33. O número esperado de recombinantes duplos é (0,01) (0,002) (100.000) = 2. Interferência = 1 – (DCO observado/DCO esperado) = 1 – 5/2 = –1,5. Por definição, a interferência é negativa.

37. a. Esse processo aparenta ser uma transdução especializada. Ela é caracterizada pela transdução de marcadores específicos com base na posição da integração do profago. Apenas aqueles genes próximos do sítio de integração são possíveis candidatos em relação à incorporação errônea em partículas de fagos, que em seguida entregam esse DNA para as bactérias receptoras.

b. Os únicos meios que ampararam o crescimento de colônias foram aqueles com ausência de cisteína ou leucina. Esses meios são selecionados para transdutantes cys^+ ou leu^+ e indicam que o profago está localizado na região *cys-leu*.

42. Não. Seria esperado que os *loci* ligados de modo próximo fossem cotransduzidos; quanto maior a frequência de cotransdução, mais próximos estão

os *loci*. Tendo em vista que apenas 1 de 858 *metE*$^+$ era também *pyrD*$^+$, os genes não estão ligados de modo próximo. A única *metE*$^+$ *pyrD*$^+$ poderia ser o resultado de cotransdução, poderia ser uma mutação espontânea de *pyrD* para *pyrD*$^+$ ou o resultado da coinfecção por dois fagos transduzidos em separado.

47. a. Para determinar quais genes estão próximos, compare as frequências de transformantes duplos. O teste par a par fornece valores baixos sempre que B está incluído, mas taxas razoavelmente altas quando qualquer outro fármaco que não B é incluído. Esse achado sugere que o gene para resistência a B não está próximo dos outros três genes.

b. Para determinar a ordem relativa dos genes para resistência a A, C e D, compare as frequências de transformantes duplos e triplos. A frequência de resistência a AC é aproximadamente a mesma da resistência a ACD, o que sugere fortemente que D está no meio. Adicionalmente, a frequência de corresistência a AD é mais alta do que a AC (sugerindo que o gene para resistência a A está mais próximo de D do que de C) e que a frequência de CD é mais alta do que de AC (sugerindo que C está mais próximo de D do que de A).

51. Para isolar as partículas de transdução especializadas do fago ϕ80 que carreavam *lac*$^+$, os pesquisadores precisariam lisogenizar a linhagem com ϕ80, induzir o fago com UV e em seguida utilizar esses lisados para transduzir uma linhagem Lac$^-$ para Lac$^+$. As colônias Lac$^+$ em seguida teriam sido utilizadas para produzir um novo lisado, o qual teria sido altamente enriquecido com o fago transdutor *lac*$^+$.

Capítulo 6

13. Com a presunção de homozigose em relação ao gene normal, o cruzamento é *A/A* · *b/b* × *a/a* · *B/B*. Os filhos seriam normais, *A/a* · *B/b*.

16. a. Vermelho

b. Roxo

c. 9 M_1/–; M_2/– roxo

3	m_1/m_1; M_2/–	azul
3	M_1/–; m_2/m_2	vermelho
1	m_1/m_1; m_2/m_2	branco

d. Os alelos mutantes não produzem enzima funcional. Entretanto, enzima funcional suficiente pode ser produzida pelo único alelo do tipo selvagem de cada gene para sintetizar níveis normais de pigmento.

20. a. O cruzamento original foi um cruzamento di-híbrido. Tanto oval quanto roxo devem representar um fenótipo de dominância incompleta.

b. Um cruzamento longo, roxo × oval, roxo é como segue:

P *L/L*; *R/R′* × *L/L′*; *R/R″*

		$\frac{1}{4}$ *R/R*	$\frac{1}{8}$ longo, vermelho
F_1	$\frac{1}{2}$ *L/L* ×	$\frac{1}{2}$ *R/R′*	$\frac{1}{8}$ longo, roxo
		$\frac{1}{4}$ *R′/R′*	$\frac{1}{8}$ longo, branco
		$\frac{1}{4}$ *R/R*	$\frac{1}{8}$ oval, vermelho
	$\frac{1}{2}$ *L′/L′* ×	$\frac{1}{2}$ *R/R′*	$\frac{1}{4}$ oval, roxo
		$\frac{1}{2}$ *R′/R′*	$\frac{1}{4}$ oval, branco

23.

Genitores	Filhos
a. AB × O	B
b. A × O	A
c. A × AB	AB
d. O × O	O

27. a. Espera-se que a proporção sexual seja de 1:1.

b. A genitora era heterozigota em relação a um alelo letal recessivo ligado ao X, que resultaria em 50% menos machos do que fêmeas.

c. Metade da progênie do sexo feminino deve ser heterozigota em relação ao alelo letal e metade deve ser homozigota em relação ao alelo não letal. Cruze individualmente as fêmeas da F_1 e determine a proporção sexual de sua progênie.

30. a. As mutações estão em dois genes diferentes, tendo em vista que o heterocário é prototrófico (as duas mutações complementaram uma à outra).

b. *leu1*$^+$; *leu2*$^-$ e *leu1*$^-$; *leu2*$^+$

c. Com a segregação independente, espere:

$\frac{1}{4}$ *leu1*$^+$; *leu2*$^-$

$\frac{1}{4}$ *leu1*$^-$; *leu2*$^+$

$\frac{1}{4}$ *leu1*$^-$; *leu2*$^-$

$\frac{1}{4}$ *leu1*$^+$; *leu2*$^+$

34. a. P *A/a* (*frizzle*) × *A/a* (*frizzle*)

F_1 1 *A/A* (normal):2 *A/a* (*frizzle*):1 *a/a* (lanosa)

b. Se *A/A* (normal) for cruzada com *a/a* (lanosa), toda a descendência será *A/a* (*frizzle*).

36. É possível a produção de descendência preta a partir de dois genitores albinos recessivos puros se o albinismo resultar de mutações em dois genes diferentes. Se o cruzamento for designado:

A/A; *b/b* × *a/a*; *B/B*

toda a descendência será:

A/a; *B/b*

e apresentará um fenótipo preto em virtude da complementação.

40. O genitor roxo pode ser *A/a*; *b/b* ou *a/a*; *B/b* para essa resposta. Presuma que o genitor roxo seja *A/a*; *b/b*. O genitor azul deve ser *A/a*; *B/b*.

43. O cruzamento é cinza × amarela ou *A/–*; *R/–* × *A/–*; *r/r*. A progênie F_1 é:

$\frac{3}{8}$ amarela $\frac{1}{8}$ preta $\frac{3}{8}$ cinza $\frac{1}{8}$ branca

Em relação à progênie branca, ambos os genitores devem carrear um alelo *r* e um *a*. Agora, o cruzamento pode ser reescrito como *A/a*; *R/r* × *A/a*; *r/r*.

46. O cão marrom original é *w/w*; *b/b* e o cão branco original é *W/W*; *B/B*. A progênie F_1 é *W/w*; *B/b* e a progênie F_2 é:

9 *W/–*; *B/–*	branca
3 *w/w*; *B/–*	preta
3 *W/–*; *b/b*	branca
1 *w/w*; *b/b*	marrom

50. Heredogramas como esse são bastante comuns. Eles indicam ausência de penetrância em virtude de epistasia ou de efeitos ambientais. O indivíduo A deve apresentar o gene autossômico dominante.

52. a. Estabeleça que W^O = Oval, W^S = Falciforme e W^R = Redonda. Os três cruzamentos são:

Cruzamento 1:	$W^S/W^S \times W^R/Y \rightarrow W^S/W^R$ e W^S/Y
Cruzamento 2:	$W^S/W^S \times W^S/Y \rightarrow W^S/W^R$ e W^R/Y
Cruzamento 3:	$W^S/W^S \times W^O/Y \rightarrow W^O/W^S$ e W^S/Y

b. $W^O/W^S \times W^R/Y$

$\frac{1}{4}$ W^O/W^R	fêmea oval
$\frac{1}{4}$ W^S/W^R	fêmea falciforme
$\frac{1}{4}$ W^O/Y	macho oval
$\frac{1}{4}$ W^S/Y	macho falciforme

55. a. Os genótipos são:

P	*B/B*; *i/i* × *b/b*; *I/I*	
F_1	*B/b*; *I/i*	sem cerdas
F_2	9 *B/–*; *I/–*	sem cerdas
	3 *B/–*; *i/i*	retas
	3 *b/b*; *I/–*	sem cerdas
	1 *b/b*; *i/i*	inclinadas

b. Os genótipos são *B/b*; *I/i* × *B/b*; *i/i*.

58. Existe um total de 159 indivíduos na progênie, que devem estar distribuídos em uma proporção de 9:3:3:1 se os dois genes estiverem se distribuindo independentemente. Você pode observar que:

Observadas	Esperadas
88 *P/–*; *Q/–*	90
32 *P/–*; *q/q*	30
25 *p/p*; *Q/–*	30
14 *p/p*; *q/q*	10

61. a. Está ocorrendo complementação, por meio da qual um produto proveniente de uma linhagem se difunde para outra linhagem e possibilita o crescimento da segunda linhagem.

b. Para que ocorra complementação, a linhagem em crescimento deve apresentar um bloqueio, que está presente mais inicialmente na via metabólica do que o bloqueio na linhagem a partir do qual a linhagem em crescimento está obtendo o produto para o crescimento.

c. Os dados sugerem que a via metabólica é:

$$trpE \rightarrow trpD \rightarrow trpB$$

d. Sem triptofano, não haveria crescimento e as células não teriam vivido durante tempo suficiente para gerar um produto que conseguisse se difundir.

63. a. A melhor explicação é que a síndrome de Marfan é herdada como um traço autossômico dominante.

b. O heredograma demonstra tanto pleiotropia (múltiplos traços afetados) quanto expressividade variável (grau variável de expressão do fenótipo).

c. A pleiotropia indica que o produto gênico é necessário em uma diversidade de tecidos, órgãos ou processos diferentes. Quando o gene é mutante, todos os tecidos que necessitam do produto do gene serão afetados. A expressividade variável de um fenótipo em relação a um determinado genótipo indica a modificação por um ou mais outros genes, influência aleatória ou efeitos ambientais.

66. a. Esse tipo de interação gênica é denominado *epistasia*. O fenótipo de *e/e* é epistático para os fenótipos de *B/–* ou *b/b*.

b. Os genótipos inferidos são como segue:

I 1 (*B/b E/e*) 2 (*B/b E/e*)

II 1 (*b/b E/e*) 2 (*B/b E/e*) 3 (*–/– e/e*) 4 (*b/b E/–*)

	5 (*B/b E/e*)	6 (*b/b E/e*)		
III	1 (*B/b E/–*)	2 (*–/b e/e*)	3 (*b/b, E/–*)	4 (*B/b E/–*)
	5 (*b/b E/–*)	6 (*B/b E/–*)	7 (*–/b e/e*)	

69. a. Uma série alélica múltipla foi detectada: superdupla > única > dupla.

b. Embora a explicação em relação à parte *a* racionalize todos os cruzamentos, ela não leva em consideração a esterilidade feminina ou a origem da planta superdupla a partir de uma variedade com flores duplas.

71. a. Um cruzamento tri-híbrido forneceria uma proporção de 63:1. Portanto, existem três *loci R* em segregação nesse cruzamento.

b.

P	R_1/R_1; R_2/R_2; R_3/R_3 × r_1/r_1; r_2/r_2; r_3/r_3		
F_1	R_1/r_1; R_2/r_2; R_3/r_3		
F_2	27	$R_1/–$; $R_2/–$; $R_3/–$	vermelho
	9	$R_1/–$; $R_2/–$; r_3/r_3	vermelho
	9	$R_1/–$; r_2/r_2; $R_3/–$	vermelho
	9	r_1/r_1; $R_2/–$; $R_3/–$	vermelho
	3	$R_1/–$; r_2/r_2; r_3/r_3	vermelho
	3	r_1/r_1; $R_2/–$; r_3/r_3	vermelho
	3	r_1/r_1; r_2/r_2; $R_3/–$	vermelho
	1	r_1/r_1; r_2/r_2; r_3/r_3	branco

c. (1) Para obter uma proporção de 1:1, apenas um dos genes pode ser heterozigoto. Um cruzamento representativo é R_1/r_1; r_2/r_2; r_3/r_3 × r_1/r_1; r_2/r_2; r_3/r_3.

(2) Para obter uma proporção de 3 vermelhos:1 branco, dois alelos devem estar segregando e não podem estar no mesmo gene. Um cruzamento representativo é R_1/r_1; R_2/r_2; r_3/r_3 × r_1/r_1; r_2/r_2; r_3/r_3.

(3) Para obter uma proporção de 7 vermelhos:1 branco, três alelos devem estar segregando e não podem estar no mesmo gene. O cruzamento é R_1/r_1; R_2/r_2; R_3/r_3 × r_1/r_1; r_2/r_2; r_3/r_3.

d. A fórmula é $1 – (1/4)n$, em que n = Número de *loci* que estão segregando nos cruzamentos representativos na parte *c*.

75. a. e **b.** A epistasia está implicada e o genótipo branco homozigoto recessivo

aparenta bloquear a produção da cor por parte de um segundo gene.

Presuma as relações de dominância a seguir: vermelha > laranja > amarela. Estabeleça que os alelos sejam designados como segue:

vermelha	A^R
laranja	A^O
amarela	A^Y

Os cruzamentos 1 a 3 agora se tornam:

P	$A^O/A^O \times A^Y/A^Y$	$A^R/A^R \times A^O/A^O$	$A^R/A^R \times A^Y/A^Y$
F_1	A^O/A^Y	A^R/A^O	A^R/A^Y
F_2	3 $A^O/–$:1 A^Y/A^Y	3 $A^R/–$:1 A^O/A^O	3 $A^R/–$:1 A^Y/A^Y

Cruzamento 4: para realizar esse cruzamento, você deve adicionar um segundo gene. Você também deve reescrever os cruzamentos 1 a 3 para incluir o segundo gene. Estabeleça que *B* possibilite a expressão da cor e que *b* bloqueie a sua expressão, produzindo branco. Os primeiros três cruzamentos se tornam:

P	A^O/A^O; $B/B \times A^Y/A^Y$; B/B
	A^R/A^R; $B/B \times A^O/A^O$; B/B
	A^R/A^R; $B/B \times A^Y/A^Y$; B/B
F_1	A^O/A^Y; B/B
	A^R/A^O; B/B
	A^R/A^Y; B/B
F_2	3 $A^O/–$; B/B:1 A^Y/A^Y; B/B
	3 $A^R/–$; B/B:1 A^O/A^O; B/B
	3 $A^R/–$; B/B:1 A^Y/A^Y; B/B

O quarto cruzamento é:

P	A^R/A^R; $B/B \times A^R/A^R$; b/b
F_1	A^R/A^R; B/b
F_2	3 A^R/A^R; $B/–$:1 A^R/A^R; b/b

Cruzamento 5: para realizar esse cruzamento, observe que não existe laranja. Portanto, os dois genitores devem carrear os alelos em relação à vermelha e à amarela, e a expressão da vermelha deve estar bloqueada.

P	A^Y/A^Y; $B/B \times A^R/A^R$; b/b	
F_1	A^R/A^Y; B/b	
F_2	9 $A^R/–$; $B/–$	vermelha
	3 $A^R/–$; b/b	branca
	3 A^Y/A^Y; $B/–$	amarela
	1 A^Y/A^Y; b/b	branca

Cruzamento 6: esse cruzamento é idêntico ao cruzamento 5, com a exceção de que a laranja substitui a amarela.

P	A^O/A^O; $B/B \times A^R/A^R$; b/b	
F_1	A^R/A^O; B/b	
F_2	9 $A^R/–$; $B/–$	vermelha
	3 $A^R/–$; b/b	branca
	3 A^O/A^O; $B/–$	laranja
	1 A^O/A^O; b/b	branca

Cruzamento 7: nesse cruzamento, a amarela é suprimida por *b/b*.

P	A^R/A^R; $B/B \times A^Y/A^Y$; b/b
F_1	A^R/A^Y; B/b
F_2	9 $A^R/–$; $B/–$ vermelha
	3 $A^R/–$; b/b branca
	3 A^Y/A^Y; $B/–$ amarela
	1 A^Y/A^Y; b/b branca

77. a. O intercruzamento de todas as linhagens mutantes que apresentam um fenótipo recessivo comum é a base do teste de complementação. Esse teste é projetado para identificar o número de genes diferentes que conseguem mutar para um fenótipo em particular. Nesse problema, se a progênie de um determinado cruzamento ainda expressar o fenótipo oscilante, as mutações falham em se complementar e são consideradas alelos do mesmo gene; se a progênie for do tipo selvagem, as mutações se complementam e as duas linhagens terão alelos mutantes de genes separados.

b. Esses dados identificam cinco grupos de complementação (genes).

c. mutante 1: $a^1/a^1 \cdot b^+/b^+ \cdot c^+/c^+ \cdot d^+/d^+ \cdot e^+/e^+$ (embora apenas alelos mutantes em geral sejam listados)

mutante 2: $a^+/a^+ \cdot b^2/b^2 \cdot c^+/c^+ \cdot d^+/d^+ \cdot e^+/e^+$

mutante 5: $a^5/a^5 \cdot b^+/b^+ \cdot c^+/c^+ \cdot d^+/d^+ \cdot e^+/e^+$

$\frac{1}{5}$ híbrido: $a^1/a^5 \cdot b^+/b^+ \cdot c^+/c^+ \cdot d^+/d^+ \cdot e^+/e^+$ fenótipo: oscilant

Conclusão: 1 e 5 são, ambos, mutantes em relação ao gene *A*.

(O cruzamento relevante $a^+/a^+ \cdot b^2/b^2 \times a^2/a^2 \cdot b^5/b^5$ fornece o híbrido a seguir.)

$\frac{2}{5}$ híbrido: $a^+/a^5 \cdot b^+/b^2 \cdot c^+/c^+ \cdot d^+/d^+ \cdot e^+/e^+$ fenótipo: tipo selvagem

Conclusão: 2 e 5 são mutantes em relação a diferentes genes.

Capítulo 7

6. A dupla-hélice do DNA é mantida unida por meio de dois tipos de ligações: covalente e de hidrogênio. As ligações covalentes são observadas dentro de cada filamento linear e ligam fortemente as bases, os açúcares e os grupos fosfato (dentro de cada componente e entre os componentes). As ligações de hidrogênio são observadas entre os dois filamentos; uma ligação de hidrogênio é formada entre uma base em um filamento e uma base no outro filamento em pareamento complementar. Essas ligações de hidrogênio são individualmente fracas, mas, coletivamente, são consideravelmente fortes.

9. Helicases são enzimas que rompem as ligações de hidrogênio que mantêm os dois filamentos de DNA unidos em uma dupla-hélice. Essa quebra é necessária para síntese de RNA e de DNA. Topoisomerases são enzimas que criam e relaxam a super-helicoidização na dupla-hélice de DNA. A própria super-helicoidização é um resultado da torção da dupla-hélice de DNA quando os dois filamentos se separam.

11. Não. A informação do DNA depende de um mecanismo de realização de cópias fiéis. As regras estritas da complementaridade asseguram que a replicação e a transcrição sejam reproduzíveis.

13. O cromossomo se tornaria irremediavelmente fragmentado.

15. b. O RNA apresentaria maior probabilidade de conter erros.

19. Se o DNA for bifilamentar, A = T, G = C e A + T + C + G = 100%. Se T = 15%, então C = [100 – 15(2)]/2 = 35%.

20. Se o DNA for bifilamentar, G = C = 24% e A = T = 26%.

24. Sim. A replicação do DNA também é semiconservativa em eucariotos diploides.

26. 5′....CCTTAAGACTAACTACTTACTGGGATC....3′

28. Sem a telomerase funcional, os telômeros seriam encurtados a cada ciclo de replicação, levando à eventual perda de informação codificadora essencial e à morte. De fato, algumas observações atuais indicam que o declínio ou a perda da atividade da telomerase desempenha um papel no mecanismo do envelhecimento em seres humanos.

30. As regras de Chargaff são A = T e G = C. Tendo em vista que essas igualdades não são observadas, a interpretação mais provável é que o DNA seja unifilamentar. O fago primeiramente deveria sintetizar um filamento complementar antes que pudesse começar a realizar múltiplas cópias de si próprio.

Capítulo 8

11. Em procariotos, a tradução inicia na extremidade 5′ enquanto a extremidade 3′ ainda está sendo sintetizada. Em eucariotos, o processamento (revestimento, recomposição) está ocorrendo na extremidade 5′ enquanto a extremidade 3′ ainda está sendo sintetizada.

17. Sim. Tanto a replicação quanto a transcrição são realizadas por máquinas moleculares grandes e multissubunidades (o replissomo e a RNA polimerase II, respectivamente), e ambas necessitam da atividade de helicase na forquilha da bolha. Entretanto, a transcrição procede em apenas um sentido e apenas um filamento de DNA é copiado.

19. a. A sequência original representa as sequências consenso –35 e –10 (com o número correto de espaços intercalares) de um promotor bacteriano. O fator σ, como parte da holoenzima RNA polimerase, reconhece e se liga a essas sequências.

b. As sequências mutadas (transpostas) não serão um sítio de ligação para o fator σ. A orientação das duas regiões, uma em relação à outra, não está correta; portanto, elas não serão reconhecidas como um promotor.

24. Íntrons autorremovíveis não são capazes de excisar a si próprios de um transcrito primário, sem a necessidade de enzimas adicionais ou de uma fonte de energia. Eles são um de muitos exemplos de moléculas de RNA que são catalíticas e, em virtude dessa propriedade, também são conhecidos como ribozimas. Com essa função adicional, o RNA é a única molécula biológica conhecida que codifica informação genética e catalisa reações biológicas. Em termos mais simples, a vida possivelmente teve início com uma molécula ou um grupo de moléculas de RNA, que desenvolveram a capacidade de se autorreplicar.

29. O RNA bifilamentar, composto por um filamento com senso (sentido) e um filamento antissenso (antissentido) complementar, pode ser utilizado por *C. elegans* (e provavelmente por todos os organismos) para prevenir seletivamente a síntese do produto gênico codificado (uma descoberta pela qual foi concedido o Prêmio Nobel de Fisiologia ou Medicina de 2006). Esse processo, denominado

silenciamento gênico, bloqueia a síntese da proteína codificada pelo gene endógeno e, portanto, é equivalente ao “nocaute” gênico. Para testar se um mRNA específico codifica uma proteína embrionária essencial, injeta-se o RNA bifilamentar produzido a partir do mRNA em zigotos ou embriões muito iniciais, ativando, assim, a via do RNAi. Os efeitos do nocaute do produto gênico especificado podem então ser acompanhados por meio da observação do que ocorre nesses embriões em comparação aos controles. Se a proteína codificada for essencial, o desenvolvimento embrionário deverá ser modificado quando o seu gene for silenciado.

Capítulo 9

13. a. e **b.** 5′ UUG GGA AGC 3′

c. e **d.** Com a presunção de que a matriz de leitura inicia na primeira base:

$$NH_3 – Leu – Gly – Ser – COOH$$

Em relação ao filamento inferior, o mRNA é 5′ GCU UCC CAA 3′ e, com a presunção de que a matriz de leitura inicia na primeira base, a cadeia de aminoácidos correspondente é:

$$NH_3 – Ala – Ser – Gln – COOH$$

17. Existem três códons em relação à isoleucina: 5′ AUU 3′, 5′ AUC 3′ e 5′ AUA 3′. Os possíveis anticódons são 3′ UAA 5′ (complementar), 3′ UAG 5′ (complementar) e 3′ UAI 5′ (oscilante). Embora seja complementar, 5′ UAU 3′ também realizaria o pareamento de bases com 5′ AUG 3′ (metionina) em virtude da oscilação e, portanto, não seria uma alternativa aceitável.

22. A estrutura quaternária ocorre em virtude de interações das subunidades de uma proteína. Nesse exemplo, a atividade enzimática que está sendo estudada pode ser aquela de uma proteína composta por duas subunidades diferentes. Os polipeptídios das subunidades são codificados por genes separados e não ligados.

26. Não. A enzima pode necessitar de modificação pós-tradução para ser ativa. As mutações nas enzimas necessárias para essas modificações não seriam mapeadas no gene da isocitrato liase.

29. Com a presunção de que todas as três mutações do gene *P* são mutações sem sentido, três possíveis códons de fim diferentes (âmbar, ocre ou opala) podem ser a causa. Uma mutação supressora seria específica para um tipo de códon sem sentido. Por exemplo, os supressores âmbar suprimiriam os mutantes âmbar, mas não os opala ou ocre.

33. As alterações de aminoácidos únicos podem resultar em alterações no dobramento das proteínas, no direcionamento proteico ou em modificações pós-tradução. Quaisquer dessas alterações dariam origem aos resultados indicados.

41. Se o anticódon em uma molécula de tRNA foi alterado por uma mutação até

apresentar o comprimento de quatro bases, com a quarta base no lado 5′ do anticódon, ele suprimiria a inserção. As alterações no ribossomo também podem induzir o deslocamento da matriz de leitura.

42. f, d, j, e, c, i, b, h, a, g.

Capítulo 10

14. A ligase é uma enzima essencial dentro de todas as células, que sela quebras no arcabouço de açúcar e fosfato do DNA. Na replicação do DNA, a ligase se une aos fragmentos de Okazaki para criar um filamento contínuo e, na clonagem, ela é utilizada para unir os diversos fragmentos de DNA ao vetor. Se ela não fosse adicionada, o vetor e o DNA clonado simplesmente desmoronariam.

15. Cada ciclo dura 5 minutos e duplica o DNA. Em 1 hora, haveria 12 ciclos; assim, o DNA seria amplificado $2^{12} = 4.096$ vezes.

18. Você poderia isolar o DNA da planta transgênica suspeita e sondar em relação à presença do transgene por meio da hibridização de *Southern*.

22. a. O fenótipo transformado será mapeado no mesmo *locus*. Se a substituição do gene ocorreu em virtude de um *crossing over* duplo, as células transformadas não conterão o DNA vetor. Se um *crossing over* único houver ocorrido, o vetor inteiro agora será parte do cromossomo linear de *Neurospora*.

b. O fenótipo transformado será mapeado para um *locus* diferente daquele do auxotrófico se o gene transformante tiver sido inserido ectopicamente (*i. e.*, em outro local). A incorporação ectópica também seria inferida por meio da PCR reversa.

23. O tamanho, as translocações entre cromossomos conhecidos e a hibridização com sondas de localização conhecida podem, todos, ser úteis na identificação de qual banda em um gel de campo pulsado corresponde a um cromossomo em particular.

33. A região do DNA que codifica a tirosinase no DNA genômico de camundongo "normal" contém dois sítios de *Eco*RI. Portanto, após a digestão com *Eco*RI, três fragmentos de tamanhos diferentes hibridizam com o clone de cDNA. Quando o DNA genômico de determinados camundongos albinos está sujeito à análise semelhante, nenhum fragmento de DNA contém sequências complementares ao mesmo cDNA. Esse resultado indica que esses camundongos apresentam ausência da capacidade de produzir tirosinase, tendo em vista que o DNA que codifica a enzima deve ter sido deletado.

36. As regiões promotora e de controle do gene de planta de interesse devem ser

clonadas e unidas na orientação correta com o gene da glicuronidase, que posiciona o gene repórter sob o mesmo controle de transcrição que o gene de interesse. O texto descreve a metodologia utilizada para criar plantas transgênicas. Transforme as células da planta com o construto do gene repórter e, conforme discutido no texto, cultive-as até plantas transgênicas. O gene da glicuronidase agora será expresso no mesmo padrão de desenvolvimento que o gene de interesse e a sua expressão pode ser facilmente monitorada ban hando-se a planta em uma solução de X-Gluc e dosando o produto azul da reação.

Capítulo 11

9. Mutantes O^C são alterações na sequência de DNA do operador que prejudicam a ligação do repressor *lac*. Tendo em vista que um operador controla apenas os genes no mesmo filamento de DNA, ele é cis (no mesmo filamento).

12. Um gene é desligado ou inativado pelo "modulador" (normalmente denominado *repressor*) no controle negativo e o repressor deve ser removido para que ocorra a transcrição. Um gene é ligado pelo "modulador" (normalmente denominado *ativador*) no controle positivo e o ativador deve ser adicionado ou convertido em uma forma ativa para que ocorra a transcrição.

21. A mutação *S* é uma alteração em *lacI*, de tal modo que a proteína repressora se liga ao operador, independentemente da presença do indutor. Em outras palavras, ela é uma mutação que inativa o sítio alostérico que se liga ao indutor, mas não afeta a capacidade do repressor de se ligar ao sítio operador. A dominância da mutação *S* ocorre em virtude da ligação do repressor mutante, até mesmo sob circunstâncias nas quais o repressor normal não se liga ao DNA (*i. e.*, na presença do indutor). As mutações reversas constitutivas que mapeiam em *lacI* são eventos mutacionais que inativam a capacidade desse repressor de se ligar ao operador. As mutações reversas constitutivas que mapeiam no operador alteram a sequência do DNA operador, de tal modo que ele não possibilitará a ligação a quaisquer moléculas repressoras (repressor do tipo selvagem ou mutante).

24. Todas as mutações em *cI*, *cII* e *cIII* afetariam a lisogenia: *cI* codifica o repressor, *cII* codifica um ativador de P_{RE} e *cIII* codifica uma proteína que protege *cII* da degradação. Mutações em *N* (um antifinalizador) também afetariam a lisogenia, tendo em vista que a sua função é necessária para a transcrição dos genes *cII* e *cIII*, mas também é necessária para os genes que apresentam papéis na lise. As mutações no gene que codifica a integrase (*int*) também afetariam a capacidade de um fago mutante de lisogenizar.

Capítulo 12

12. Em geral, o estado fundamental de um gene bacteriano é “ligado”. Portanto, o início da transcrição é evitado ou reduzido se a ligação da RNA polimerase estiver bloqueada. Contrariamente, o estado fundamental de eucariotos é “desligado”. Portanto, o maquinário de transcrição (incluindo a RNA polimerase II e os fatores de transcrição gerais correlatos) não consegue se ligar ao promotor na ausência de outras proteínas reguladoras.

16. Entre as mutações que poderiam prevenir que uma linhagem de levedura realizasse a alteração de seu tipo reprodutivo, estariam as mutações nos genes *HO* e *HMRa*. O gene *HO* codifica uma endonuclease que corta o DNA para iniciar a alteração e o *locus HMRa* contém o “cassete” de informação genética não expressa para o tipo reprodutivo MATa.

19. O termo *herança epigenética* é utilizado para descrever as alterações herdáveis nas quais a própria sequência de DNA não é alterada. Ele pode ser definido operacionalmente como a herança de estados da cromatina de uma geração celular até a próxima. O *imprinting* genômico, a inativação do cromossomo X e a variegação por efeito de posição são alguns exemplos.

23. Acredita-se que a herança da estrutura da cromatina seja responsável pela herança da informação epigenética. Essa herança ocorre em virtude da herança do código de histonas e também pode incluir a herança dos padrões de metilação do DNA.

36. Um gene não expresso em virtude da alteração da sua sequência de DNA nunca será expresso e será herdado de geração em geração. Um gene inativado epigeneticamente ainda pode ser regulado. A estrutura da cromatina pode ser alterada no decorrer do ciclo celular; por exemplo, quando fatores de transcrição modificam o código de histonas.

38. A estrutura da cromatina afeta amplamente a expressão gênica. Os transgenes inseridos em regiões de eucromatina apresentariam mais probabilidade de serem capazes de expressar-se do que aqueles inseridos em regiões de heterocromatina.

Capítulo 13

12. O gene de regra dos pares primário *eve* (*even-skipped*) seria expresso em sete listras ao longo do eixo A-P do blastoderma tardio.

15. Se você diagramar esses resultados, verificará que a deleção de um gene que atua posteriormente possibilita que os segmentos próximos mais anteriores estendam-se em sentido posterior. A deleção de um gene anterior não possibilita a extensão do segmento próximo mais posterior em um sentido anterior. Os genes *gap* ativam o *Ubx* tanto no segmento torácico quanto no abdominal, enquanto os genes *abd-A* e *Abd-B* são ativados apenas nos segmentos abdominal intermediário e posterior. O funcionamento dos genes *abd-A* e *Abd-B* naqueles segmentos de algum modo evita a expressão de *Ubx*. Entretanto, se os genes *abd-A* e *Abd-B* forem deletados, *Ubx* poderá ser expresso nessas regiões.

18. a. Um gene de regra dos pares.

b. Observe a expressão do mRNA do gene candidato em um padrão repetido de sete listras ao longo do eixo A-P do embrião em desenvolvimento.

c. Não. Um embrião mutante em relação ao gene *gap Krüppel* não teria muitos segmentos anteriores. Esse efeito seria epistático sobre a expressão de um gene de regra dos pares.

21. a. O homeodomínio é um domínio proteico conservado, que contém 60 aminoácidos, observado em um número significativo de fatores de transcrição. Qualquer proteína que contém um homeodomínio funcional é quase certamente um fator de transcrição de ligação a uma sequência específica de DNA.

b. O gene *eyeless* (denominado por seu fenótipo mutante) regula o desenvolvimento dos olhos em *Drosophila*. Você esperaria que ele fosse expresso apenas naquelas células que darão origem aos olhos. Para testar essa previsão, a localização da expressão do mRNA de *eyeless* deve ser visualizada utilizando-se hibridização *in situ* e a da proteína Eyeless utilizando-se métodos imunológicos. Por meio da manipulação genética, o gene *eyeless* pode ser expresso em tecidos nos quais ele não é ordinariamente expresso. Por exemplo, quando o *eyeless* é ativado em células que se destinam a formar as pernas, olhos são formados nas pernas.

c. Experimentos transgênicos demonstraram que o gene *Small eye* de camundongo e o gene *eyeless* de *Drosophila* são tão semelhantes que o gene de camundongos pode substituir o *eyeless* quando introduzido em *Drosophila*. Assim como na resposta da parte *b*, quando o gene *Small eye* de camundongo é expresso em *Drosophila*, até mesmo em células que se destinam a formar pernas, olhos são formados nas pernas. (Entretanto, os "olhos" não são olhos de camundongo, tendo em vista que o *Small eye* e o *eyeless* atuam como interruptores máster que ligam toda a cascata de genes necessária para a formação do olho – nesse caso, o conjunto de *Drosophila* para formar um olho de *Drosophila*.)

25. A proteína GLP-1 está localizada nas duas células anteriores do embrião de quatro células de *C. elegans* por meio da repressão de sua tradução nas duas células posteriores. A repressão da tradução de GLP-1 requer a região de controle espacial 3′ UTR (SCR). A deleção da SCR possibilitará a expressão de *glp-1* tanto em células anteriores quanto posteriores. Tanto em mutantes heterozigotos quanto em homozigotos, você esperaria a expressão da proteína GLP-1 em todas as células.

Capítulo 14

11. Tendo em vista que as bactérias apresentam genomas pequenos (aproximadamente 3 Mb) e essencialmente nenhuma sequência repetida, a abordagem de *shotgun* do genoma inteiro seria utilizada.

13. Um arcabouço também é denominado um supercontig. Um contig é uma sequência de leituras sobrepostas montadas em uma unidade e um arcabouço é uma coleção de contigs unidos.

18. Sim. O operador é o local no qual o repressor se liga funcionalmente por meio de interações da sequência de DNA com a proteína repressora.

23. Você pode determinar se o clone de cDNA é ou não um monstro por meio do alinhamento da sequência de cDNA com a sequência genômica. (Estão disponíveis programas de computador para a realização dos referidos alinhamentos.) A sequência é derivada de dois sítios diferentes? O cDNA mapeia em uma região (do tamanho do gene) no genoma ou em duas regiões diferentes? Os íntrons podem complicar a questão.

27. a. Tendo em vista que o código triplo é redundante, podem ocorrer alterações na sequência de nucleotídios do DNA (especialmente naqueles nucleotídios que codificam a terceira posição de um códon) sem alteração da proteína codificada.

b. Pode-se esperar que as sequências proteicas evoluam e divirjam mais lentamente do que os genes que as codificam.

33. A montagem correta de regiões grandes e quase idênticas é problemática com qualquer método de sequenciamento genômico. Entretanto, o método *shotgun* de genoma inteiro é menos eficaz para encontrar essas regiões do que o método com base em clone. Esse método também apresenta a vantagem adicional do fácil acesso ao(s) clone(s) suspeito(s) para a análise adicional.

36. Quinze por cento são funções gênicas essenciais (tais como as enzimas necessárias para a replicação do DNA ou a síntese de proteínas).

Vinte e cinco por cento são auxotróficos (enzimas necessárias para a síntese de aminoácidos ou para o metabolismo de açúcares etc.).

Sessenta por cento são redundantes, ou vias não testadas (genes para histonas, tubulina, RNA ribossômicos etc. estão presentes em múltiplas cópias; a levedura

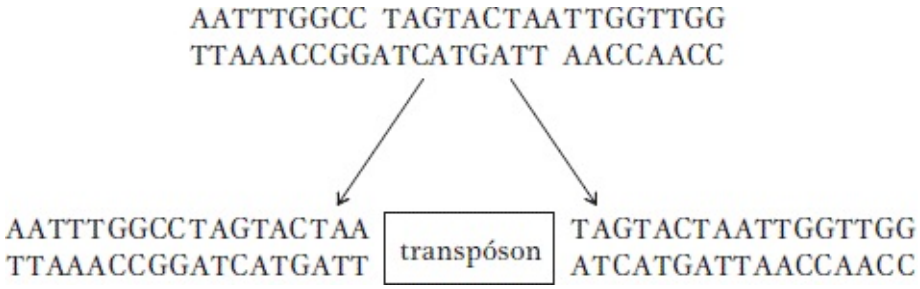
pode necessitar de muitos genes apenas em situações únicas, especiais ou de outros modos que não são necessários para a vida em laboratório).

Capítulo 15

9. Boeke, Fink e seus colaboradores demonstraram que a transposição do elemento *Ty* em leveduras ocorre por meio de um RNA intermediário. Eles construíram um plasmídio com a utilização de um elemento *Ty*, no qual inseriram não apenas um promotor que pode ser ativado por galactose, mas também um íntron na região codificadora do elemento *Ty*. Primeiramente, a frequência de transposição foi amplamente aumentada por meio da adição de galactose, que indicou que um aumento na transcrição (e na produção de RNA) estava correlacionado às taxas de transposição. Mais importante, após a transposição, eles observaram que o DNA de *Ty* recém-transposto não apresentava sequência do íntron. Tendo em vista que a remoção de íntron ocorre apenas durante o processamento do RNA, deve ter havido um RNA intermediário no evento de transposição.

13. Alguns elementos de transposição desenvolveram estratégias para se inserir em abrigos seguros, regiões do genoma onde eles causarão danos mínimos. Os abrigos seguros incluem genes duplicados (tais como genes de tRNA ou rRNA) e outros elementos de transposição. Os abrigos seguros em genomas bacterianos podem ser sequências muito específicas entre os genes ou genes de rRNA repetidos.

27. O corte desencontrado levará a uma duplicação de nove pares de bases no sítio-alvo que flanqueia o transpóson inserido.



30. Não seria uma surpresa observar um elemento SINE no íntron de um gene, em vez de em um éxon. O processamento do pré-mRNA removeria o elemento de transposição como parte do íntron e a tradução da enzima FB não seria efetuada.

Capítulo 16

8. Você precisa conhecer a matriz de leitura da possível mensagem.

11. Com a presunção de substituições de pares de bases únicas, CGG pode ser alterado para CGU, CGA, CGC ou AGG e ainda codificará a arginina.

14. A lista de observações a seguir argumenta que "o câncer é uma doença genética":

(1) Determinados cânceres são herdados como traços mendelianos simples altamente penetrantes.

(2) A maior parte dos agentes carcinogênicos também é mutagênica.

(3) Diversos oncogenes foram isolados a partir de vírus tumorais.

(4) Uma diversidade de genes que levam à suscetibilidade a tipos de câncer em particular foi mapeada, isolada e estudada.

(5) Foram isolados oncogenes dominantes de células tumorais.

(6) Determinados cânceres estão altamente correlacionados com rearranjos cromossômicos específicos.

17. O T malpareado seria corrigido para C e o ACG resultante, após a transcrição, seria 5′ UGC 3′ e codificaria a cisteína. Ou, se o outro filamento fosse corrigido, ATG seria transcrito para 5′ UAC 3′ e codificaria a tirosina.

24. Estão disponíveis muitos sistemas de reparo: reversão direta, reparo por excisão, reparo acoplado à transcrição e junção de extremidades não homólogas.

25. Sim, ele é mutagênico. Ele causará transições de CG para TA.

35. a. Uma ausência de revertentes sugere uma deleção ou uma inversão no gene.

b. Para compreender esses dados, relembre que metade da progênie deve ser originária do genitor do tipo selvagem.

Prototrófico A: tendo em vista que 100% da progênie é prototrófica, deve ter ocorrido uma reversão no sítio mutante original.

Prototrófico B: metade da progênie é de prototróficos parentais e, dos prototróficos remanescentes, 28%, são resultantes de nova mutação. Observe que 28% são aproximadamente iguais aos 22% de auxotróficos. A sugestão é de que tenha ocorrido uma mutação supressora não ligada, produzindo a segregação independente com o mutante *nic-2*.

Prototrófico C: existem 496 prototróficos “revertentes” (os outros 500 são prototróficos parentais) e quatro auxotróficos. Isso sugere que uma mutação supressora tenha acontecido em um sítio muito próximo da mutação original e raramente tenha sido separada da mutação original por meio de recombinação [100% (4 × 2)/1.000 = 0,8 u.m.].

38. O xeroderma pigmentoso é um distúrbio genético heterogêneo causado por mutações em qualquer um de diversos genes que participam no processo de NER (reparo por excisão de nucleotídio). Assim como atesta a descoberta de outra proteína na via NHEJ por meio de pesquisas na linhagem celular 2BN, esse paciente pode apresentar uma mutação em um gene ainda desconhecido que codifique uma proteína necessária para NER.

Capítulo 17

21. MM N OO seria classificado como 2*n* – 1, MM NN OO seria classificado como 2*n* e MMM NN PP seria classificado como 2*n* + 1.

24. Haveria um possível quadrivalente.

27. Sete cromossomos.

29. As células destinadas a se tornar grãos de pólen podem ser induzidas por meio de tratamento a frio para o crescimento como embrioides. Esses embrioides, em seguida, podem ser cultivados em ágar para formar mudas monoploides.

31. Sim.

34. Não.

36. Um fragmento acêntrico não pode ser alinhado ou movimentado na meiose (ou na mitose) e, consequentemente, é perdido.

39. Deleções muito grandes tendem a ser letais, provavelmente em virtude do desequilíbrio genômico ou da exposição de genes letais recessivos. Portanto, a alça de pareamento muito grande observada mais provavelmente é originária de uma inversão heterozigota.

41. A síndrome de Williams é o resultado de uma deleção da região 7q11.23 do cromossomo 7. A síndrome *cri du chat* é o resultado de uma deleção de uma parte significativa do braço curto do cromossomo 5 (especificamente das bandas 5p15.2 e 5p15.3). Tanto a síndrome de Turner (XO) quanto a síndrome de Down (trissomia do cromossomo 21) resultam da não disjunção meiótica. O termo *síndrome* é utilizado para descrever um conjunto de fenótipos (com frequência complexos e variados) que em geral estão presentes juntos.

46. A ordem é *b a c e d f*:

Alelo	Banda
b	1
a	2
c	3
e	4
d	5

47. Os dados sugerem que um dos pontos de quebra da inversão, ou ambos, estão localizados em um gene essencial, causando uma mutação letal recessiva.

50. a. Quando cruzados com fêmeas amarelas, os resultados seriam:

X^e/Y^{e+}	machos cinza
X^e/X^e	fêmeas amarelas

b. Se o alelo e^+ fosse translocado para um autossomo, a progênie seria como segue, em que "A" indica autossomo:

P	$A^{e+}/A; X^e/Y \times A/A; X^e/X^e$
F_1	$A^{e+}/A; X^e/X^e$ fêmea cinza
	$A^{e+}/A; X^e/Y$ macho cinza
	$A/A; X^e/X^e$ fêmea amarela
	$A/A; X^e/Y$ macho amarelo

52.

Síndrome de Klinefelter	homem XXY
Síndrome de Down	trissomia do cromossomo 21
Síndrome de Turner	mulher XO

56. a. Se um hexaploide fosse cruzado com um tetraploide, o resultado seria um pentaploide.

b. O cruzamento de *A/A* com *a/a/a/a* para obter *A/a/a*.

c. O modo mais fácil é expor as células da planta *A/a** à colchicina durante uma divisão celular, o que resultará em uma duplicação dos cromossomos para produzir *A/A/a*/a**.

d. Cruze um hexaploide (*a/a/a/a/a/a*) com um diploide (*A/A*) para obter *A/a/a/a*.

58. a. A proporção de plantas com folhas normais e com folhas *potato* será de 5:1.

b. Se o gene não estiver no cromossomo 6, deverá haver uma proporção de 1:1 de plantas com folhas normais e com folhas *potato*.

62. a. A planta aberrante é semiestéril, o que sugere uma inversão. Tendo em vista

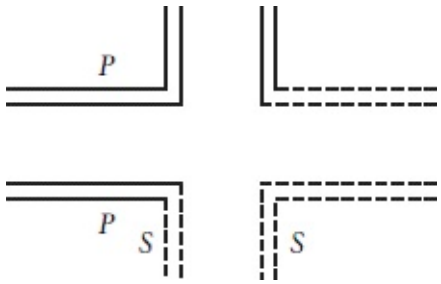
que as frequências de recombinação *d-f* e *y-p* na planta aberrante são normais, a inversão deve implicar *b* a *x*.

b. A obtenção de progênie recombinante quando houve uma inversão requer a ocorrência de um *crossing over* duplo dentro da região invertida ou de *crossovers* únicos entre *f* e a inversão, que ocorreu em algum lugar entre *f* e *b*.

64. A planta original é homozigota em relação a uma translocação entre os cromossomos 1 e 5, com os pontos de quebra muito próximos dos genes *P* e *S*. Em virtude da ligação próxima, foi observada uma proporção que sugere um cruzamento mono-híbrido, em vez de um cruzamento di-híbrido, tanto com autopolinização quanto com cruzamento-teste. Todos os gametas são férteis, em virtude da homozigose.

planta original:	*P S/p s*
testador:	*p s/p s*

Progênie F_1: heterozigota em relação à translocação:



O modo mais fácil de testar essa hipótese é observar os cromossomos dos heterozigotos na meiose I.

70. Os genitores originais devem ter apresentado a constituição cromossômica a seguir:

G. hirsutum	26 grandes, 26 pequenos
G. thurberi	26 pequenos
G. herbaceum	26 grandes

O *G. hirsutum* é um poliploide derivado de um cruzamento entre as duas espécies do Velho Mundo, o que pode ser facilmente verificado por meio da observação dos cromossomos.

72. a. Perda de um X no feto em desenvolvimento após o estágio de duas células.

b. Não disjunção que leva à síndrome de Klinefelter (XXY), seguida por um evento de não disjunção em uma célula envolvendo o cromossomo Y após o estágio de duas células, resultando em XX e XXYY.

c. Não disjunção do X no estágio de uma célula.

d. Zigotos XX e XY fundidos (a partir de fertilizações em separado de dois ovócitos ou de um ovócito e um corpúsculo polar por um espermatozoide contendo um X e outro contendo um Y).

e. Não disjunção do X no estágio de duas células ou posteriormente.

75. a. Cada mutante é cruzado com o tipo selvagem, ou:

$$m \times m^+$$

As tétrades (óctades) resultantes demonstram segregação de 1:1, indicando que cada mutante é o resultado de uma mutação em um único gene.

b. Os resultados do cruzamento das duas linhagens mutantes indicam que ambas as linhagens são mutantes em relação ao mesmo gene:

$$m_1 \times m_2$$

ou que elas são mutantes em genes diferentes, porém estreitamente ligados:

$$m_1\ m_2^+ \times m_1^+\ m_2$$

c. e **d.** Tendo em vista que a progênie fenotipicamente preta pode resultar de não disjunção (observe que, nos casos C e D, preto aparece em conjunto com esporos abortados), o mutante 1 e o mutante 2 provavelmente são mutantes em genes diferentes, porém estreitamente ligados. Portanto, o cruzamento é:

$$m_1\ m_2^+ \times m_1^+\ m_2$$

O caso A é uma tétrade DNP e seria o resultado de um *crossing over* duplo de quatro filamentos.

$m_1^+\ m_2^+$	preto
$m_1^+\ m_2^+$	preto
$m_1\ m_2$	castanho-claro
$m_1\ m_2$	castanho-claro

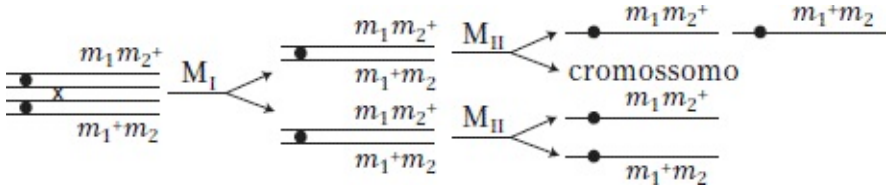
O caso B é um tetratipo e seria o resultado de um *crossover* único entre um dos genes e o centrômero.

$m_1^+ m_2^+$	preto
$m_1^+ m_2$	castanho-claro
$m_1 m_2^+$	castanho-claro
$m_1 m_2$	castanho-claro

O caso C é o resultado da não disjunção na meiose I.

$m_1^+ m_2^+$; $m_1^+ m_2^+$	preto
$m_1^+ m_2^+$; $m_1^+ m_2^+$	preto
nenhum cromossomo	aborto
nenhum cromossomo	aborto

O caso D é o resultado de recombinação entre um dos genes e o centrômero, seguida por não disjunção na meiose II. Por exemplo:



$m_1^+ m_2$; $m_1 m_2^+$	preto
nenhum cromossomo	aborto
$m_1 m_2^+$	castanho-claro
$m_1^+ m_2$	castanho-claro

Capítulo 18

7. A frequência de um alelo em uma população pode ser alterada por meio de seleção natural, mutação, migração e deriva genética.

11. A frequência de *b* é $q = \sqrt{0{,}04} = 0{,}2$ e a frequência de *B* é $p = 1 - q = 0{,}8$. A frequência de *B/B* é $p^2 = 0{,}64$ e a frequência de *B/b* é $2pq = 0{,}32$.

14. a. $p' = 0{,}5\,[(0{,}5)\,(1{,}0) + 0{,}5\,(1{,}0)]/[(0{,}25)\,(1{,}0) + (0{,}5)\,(1{,}0) + (0{,}25)\,(0{,}7)] = 0{,}54$

b. 0,008

13. $\frac{1}{2}(0{,}02) + \frac{1}{2}(0{,}02)^2 = 0{,}0102$

21. probabilidade de fixação $= \frac{1}{2N} = \frac{1}{100.000}$, probabilidade de perda $= 1 - \frac{1}{2N} = \frac{99.999}{100.000}$

26. a. $F_I = (\frac{1}{2})^3 \times (1 + \frac{1}{2}) = 3/16$

b. $\frac{1}{8} = (\frac{1}{2})^3 \times (1 + F_A)$, assim $F_A = 0$

29. $p_A = p_a = p_B = p_b = 0{,}5$. Em equilíbrio, a frequência de indivíduos duplamente heterozigotos é $2\,(p_A p_a) \times 2\,(p_B p_b) = 0{,}25$.

31. Antes da migração, $q_A = 0{,}1$ e $q_B = 0{,}3$ nas duas populações. Tendo em vista que as duas populações são iguais em número, imediatamente após a migração $q_{A+B} = \frac{1}{2}\,(q_A + q_B) = \frac{1}{2}\,(0{,}1 + 0{,}3) = 0{,}2$. No novo equilíbrio, a frequência de homens afetados é $q = 0{,}2$ e, então, a frequência de mulheres afetadas é $q^2 = (0{,}2)^2 = 0{,}04$. (O daltonismo é um traço ligado ao X.)

33. $q^2 = 0{,}002$, $q = 0{,}045$. Presumindo que *F* nos fundadores seja 0,0, $F_{50} = 0{,}222$ (ver Quadro 18.3). $f_{a/a} = q^2 + pqF = 0{,}012$.

37. $\hat{H} = [4 \times 50.000 \times (3 \times 10^{-8})]/[4 \times 50.000 \times (3 \times 10^{-8}) + 1] = 5{,}96 \times 10^{-3}$

41. a. $\hat{q} = \sqrt{1{,}0 \times 10^{-5}/0{,}5} = 4{,}47 \times 10^{-3}$

Custo genético $= sq^2 = 0{,}5\,(4{,}47 \times 10^{-3})^2 = 1^{-5}$

b. $\hat{q} = 6{,}32 \times 10^{-3}$

Custo genético $= sq^2 = 0{,}5\,(6{,}32 \times 10^{-3})^2 = 2 \times 10^{-5}$

c. Custo genético $= sq^2 = 0{,}3\,(5{,}77 \times 10^{-3})^2 = 10^{-5}$

Capítulo 19

7. Muitos traços variam mais ou menos continuamente ao longo de uma ampla gama. Por exemplo, a altura, o peso, a forma, a cor, a taxa reprodutiva, a atividade metabólica etc. variam quantitativa, não qualitativamente. A variação contínua com frequência pode ser representada por uma curva sinusoidal, na qual o fenótipo "médio" é mais comum do que os extremos. A variação descontínua descreve os fenótipos discretos e facilmente classificáveis da genética mendeliana simples: o formato da semente, os mutantes auxotróficos, a anemia falciforme etc. Esses traços com frequência demonstram uma relação simples entre o genótipo e o fenótipo, embora traços descontínuos tais como afetado *versus* não afetado em relação a uma condição de doença também possam exibir herança complexa.

9. A média é de 4,7 cerdas, a variância é de 1,11 cerdas² e o desvio padrão é de 1,05 cerda.

12. O cultivador não pode ter certeza de que essa população responderá ao cruzamento seletivo, muito embora a herdabilidade no sentido amplo seja alta. A herdabilidade no sentido amplo é a proporção da variância genética em relação à fenotípica. A variância genética é a soma das variâncias aditiva e de dominância. Apenas a variância aditiva é transmitida do genitor para a descendência. A variância de dominância não é transmitida do genitor para a descendência. Se toda a variância genética na população for de dominância, então o cruzamento seletivo não será bem-sucedido.

18. $\bar{x}_{par}$ = [(9,8 + 10,8)/2] – 9,6 = 0,7 mm, $\bar{a}_{par}$ =0,79 × 0,7 = 0,55 mm, $\hat{x}_{off}$ = 9,6 + 0,55 = 10,15 mm

23. V_e = 3,5 g², V_g na população B é 21,0 – 3,5 = 17,5 g², H^2 = 17,5/21,0 = 0,83.

Capítulo 20

9. Os três princípios são (1) os indivíduos de qualquer população variam uns dos outros, (2) a descendência se assemelha aos seus genitores mais do que se assemelha aos indivíduos não relacionados e (3) alguns tipos obtêm mais sucesso na sobrevivência e na reprodução do que outros tipos em um determinado ambiente.

11. A taxa relativa de substituições sinônimas e não sinônimas não seria mais alta do que a esperada em um pseudogene de globina, tendo em vista que um pseudogene é inativo e não apresenta função a ser preservada.

13. Uma população não será diferenciada de outras por meio de autocruzamento local se:

$$\mu \geq 1/N$$

e, assim:

$$N \geq 1/\mu$$

$$N \geq 105$$

17. Quando alterações de aminoácidos forem direcionadas pela seleção adaptativa positiva, deve haver um excesso de alterações não sinônimas. O gene *MC1R* (receptor de melanocortina 1) codifica uma proteína-chave que controla a quantidade de melanina na pele e nos cabelos. As populações asiáticas e europeias aparentam ter vivenciado a seleção adaptativa positiva em relação à pele menos pigmentada, em relação às suas correspondentes africanas.

19. Sequências não codificadoras. Uma importante restrição na evolução gênica compreende os possíveis efeitos pleiotrópicos de mutações em regiões codificadoras. Esses efeitos podem ser contornados por mutações em sequências reguladoras, que desempenham um papel importante na evolução da forma corporal. As alterações nas sequências não codificadoras proporcionam um mecanismo para alterar um aspecto da expressão gênica, enquanto preservam o papel das proteínas pleiotrópicas em outros processos essenciais do desenvolvimento.

21. a. A mutação HbS surgiu independentemente em cinco haplótipos diferentes

em regiões diferentes e em seguida aumentou até uma alta frequência.

c. Duas linhagens independentes de bacteriófagos desenvolveram a capacidade de se reproduzir em altas temperaturas em um novo hospedeiro.

23. Um novo gene originado por duplicação pode (1) desenvolver uma nova função, (2) tornar-se inativado ou (3) realizar parte da função original, compartilhando a função total com o gene original.

25. Para sítios polimórficos com uma espécie, estabeleça que não sinônimo = *a* e sinônimo = *b*. Para os sítios polimórficos entre as espécies, estabeleça que não sinônimo = *c* e sinônimo = *d*. Se a divergência ocorrer em virtude da evolução neutra, então:

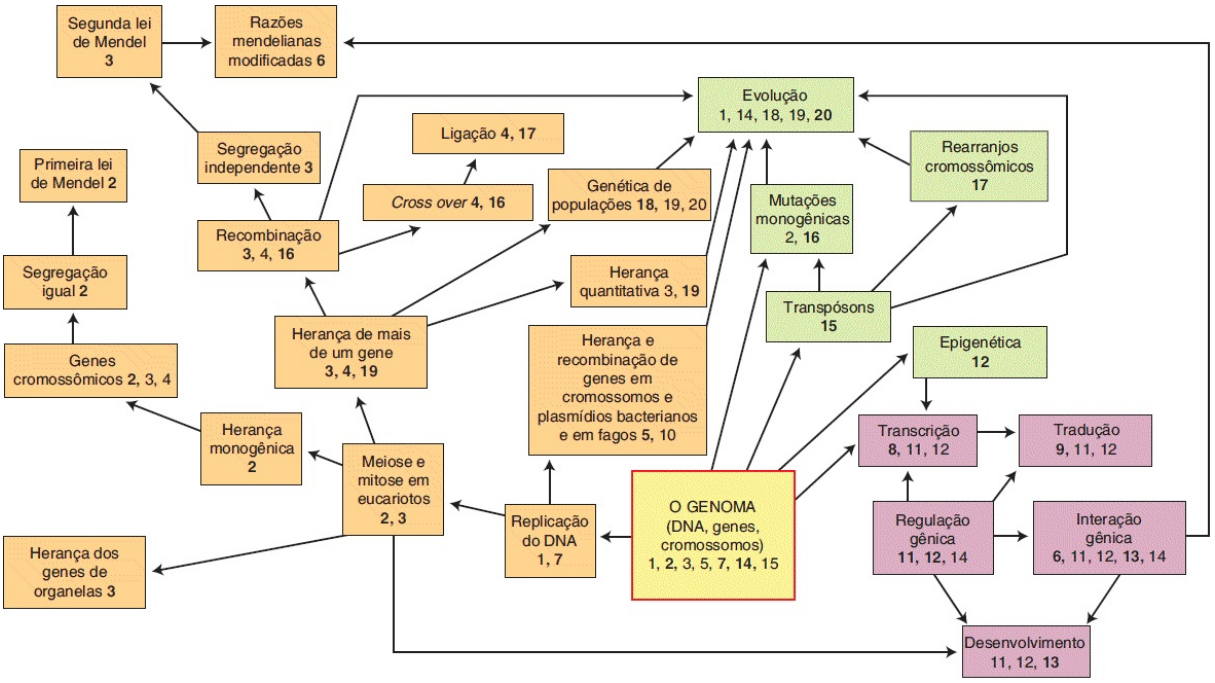
$$a/b = c/d$$

Se a divergência ocorrer em virtude da seleção, então:

$$a/b < c/d$$

Entretanto, nesse exemplo, $a/b = 20/50 > c/d = 2/18$, que não corresponde a nenhuma expectativa. Tendo em vista que a proporção de polimorfismos não sinônimos e sinônimos (a/b) é relativamente alta, o gene que está sendo estudado pode codificar uma proteína tolerante de diferenças relativamente menores entre espécies. As diferenças relativamente menores entre espécies podem sugerir que a especiação foi um evento recente, de modo que novos polimorfismos foram fixados em uma espécie que não são variantes na outra.

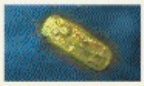
Mapa da Genética



Mapa das divisões gerais da genética em boxes, com as principais conexões entre eles tratadas neste livro ligadas por setas. Em geral, laranja diz respeito à herança; roxo à função; e verde à mudança. Os números representam os capítulos que versam sobre cada tópico, sendo grafados em negrito aqueles com as principais discussões.

Índice de Organismos-modelo

A tabela a seguir fornece referências das páginas para discussões de organismos-modelo específicos no texto.

	 Bactéria (*E. coli*)	Levedura (*S. cerevisiae*)	Mofo do pão (*N. crassa*)	Mostarda (*A. thaliana*)	Nematódeo (*C. elegans*)	Mosca-das-frutas (*D. melanogaster*)	Camundongo (*M. musculus*)
Características principais	p. 698	p. 700	p. 702	p. 704	p. 706	p. 708	p. 710
Destaques	p. 160	p. 384	p. 92		p. 435	p. 48, pp. 416-417	p. 197
Capítulos							
1. A Revolução Genética	descrição do organismo, p. 11	descrição do organismo, p. 11	experimento "um gene-uma enzima" de Beadle e Tatum, p. 7	descrição do organismo, p. 11	descrição do organismo, p. 11	descrição do organismo, p. 11	descrição do organismo, p. 11 estoques de linhagens de camundongos, p. 11
2. Herança Monogênica		análise genética usando asco, p. 38	mutantes de desenvolvimento do micélio, p. 28	mutantes de desenvolvimento da flor, pp. 28-29		identificação de um gene para desenvolvimento da asa, pp. 44-45 determinação do sexo, pp. 46-50	
3. Segregação Independente de Genes		recombinação meiótica, pp. 94, 95	ciclo de vida, p. 90 observação de segregação independente, p. 91 herança materna (mutantes *poky*), p. 98 segregação citoplasmática, pp. 98-100			herança ligada ao X (cor do olho), pp. 91-92	
4. Mapeamento de Cromossomos Eucarióticos por Recombinação		análise de tétrades de *crossovers*, p. 118	mapeamento de centrômero, pp. 129-131			mapa do cromossomo de *Drosophila*, p. 113 experimentos de Morgan, ligação, pp. 115-116 análise de ligação di-híbrida, pp. 119-121 cruzamento-teste de três pontos, pp. 121-123 interferência, p. 124 não há *crossover* em machos, p. 124	
5. A Genética das Bactérias e seus Vírus	conjugação, pp. 157-158 cruzamentos de fagos, pp. 172-173 mapeamento do gene *rII*, p. 173 experimentos clássicos de transdução, pp. 173-178 mapa genômico, pp. 178-180						
6. Interação Gênica		supressão, pp. 207-208 mutações modificadoras, p. 210	experimentos de Beadle-Tatum, pp. 199-200 complementação, pp. 203-204			supressão da cor do olho, p. 208	exemplo de gene haploinsuficiente, p. 193 coloração da pelagem, pp. 196-197
7. DNA | Estrutura e Replicação	experimento de Hershey-Chase, p. 232 experimento de Meselson-Stahl, pp. 238-240 DNA polimerases, p. 241 velocidade de replicação, p. 244 replissomo, pp. 244-245 origem da replicação, p. 247	replissomo, p. 247 origens de replicação, p. 247 controle do ciclo celular, p. 248 complexo de reconhecimento da origem, p. 242					experimento de Griffith, pp. 230-231
8. RNA | Transcrição e Processamento	número de genes, p. 258 experimento de pulso e caça de Volkin-Astrachan, p. 259 estágios da transcrição, pp. 263-266 sequências promotoras, pp. 263-264 fatores sigma, p. 264 término, p. 265 densidade gênica, p. 266	número de genes, p. 258 RNA polimerase II, p. 267 frequência de íntrons, p. 270				número de genes, p. 258 densidade gênica, p. 266	
9. Proteínas e sua Síntese	experimento do comprimento do códon de Crick, p. 288	abundância de transcritos de RNA, p. 285 elucidação do código genético de Nirenberg, p. 290		número de genes de quinase, pp. 302-303	interatoma, p. 301		
10. Isolamento e Manipulação de Genes	enzimas de restrição, pp. 314, 318-319 clonagem com vetores bacteriófagos, pp. 321-322 clonagem com fosmídios, p. 322	engenharia genética com vetores de levedura, p. 337			tamanho do genoma e clonagem, p. 324 transgênese em, p. 340		transgênese em, pp. 340-341 nocaute de gene alvo, p. 341
11. Regulação da Expressão Gênica em Bactérias e seus Vírus	experimentos do óperon *lac* de Jacob-Monod, pp. 350-361 óperon *trp*, pp. 364-366						
12. Regulação da Expressão Gênica em Eucariotos		sistema GAL, pp. 383-387 mutantes SWI-SNF, p. 391 modificação das histonas, pp. 391-394 controle do tipo reprodutivo, p. 388 silenciamento gênico e mudança do tipo reprodutivo, pp. 399-400				variegação de efeito de posição, pp. 400-402	*imprinting* genômico, pp. 403-404 compensação de dose, pp. 405-406
13. Controle Genético do Desenvolvimento					como organismo-modelo, p. 417 destinos das linhagens celulares, pp. 435-438 tempo de desenvolvimento, p. 438	mutantes homeóticos, p. 414 como organismos-modelo, pp. 416-417 genes homeóticos, pp. 418-421 desenvolvimento inicial, pp. 418-421 determinação do sexo, p. 434 papéis múltiplos do gene *hedgehog*, pp. 438-439	como organismo-modelo, p. 414 aglomerados do gene *Hox*, p. 422

14. Genomas e Genômica	genômica comparativa de patogênicos e não patogênicos, pp. 468-470 mutagênese sistemática direcionada, p. 474	ano de sequenciamento, p. 448 preenchimento de espaços na sequência, p. 456 teste di-híbrido, p. 472			ano de sequenciamento, p. 449	ano de sequenciamento, p. 449 método de sequenciamento, p. 456 repete em, p. 456 tendenciosidade de códon, pp. 432, 459	genômica comparativa humano–camundongo, pp. 466-467 uso na identificação de elementos codificadores conservados, pp. 462-464
15. Genoma Dinâmico	Elementos de Transposição	elementos IS, pp. 486-487	elementos *Ty*, pp. 489-493 falta de transposons de DNA, p. 495 direcionamento de transposon, p. 500				elementos tipo *copia*, p. 492 elementos *P*, pp. 493-494 disgenesia híbrida, pp. 493-494 elemento *P* para marcação/transgênese, p. 496 direcionamento de transpóson, p. 500	vírus do tumor mamário de camundongo, p. 490
16. Mutação, Reparo e Recombinação	teste da flutuação de Luria-Dellbrück, pp. 515-518 metiltransferase, p. 525 pontos quentes de mutação, pp. 526-527 reparo de mal pareamento, pp. 528-530 reparo SOS, p. 531	número de quebras no filamento duplo, p. 535	recombinação meiótica, p. 533				
17. Alterações Cromossômicas em Grande Escala		duplicação do genoma inteiro, pp. 563-564			cromossomos balanceadores, p. 568	poliploides experimentais, p. 552 não disjunção, p. 553 aneuploidia, p. 545 compensação de dose, p. 558 pseudodominância e mapeamento de deleção, p. 561 inversões, pp. 564-565 cromossomos balanceadores, p. 568 variegação de efeito de posição, p. 570	
18. Genética de Populações	taxa de mutação, p. 606	diversidade nucleotídica, p. 604 taxa de mutação, p. 606	diversidade nucleotídica, p. 604	diversidade nucleotídica, p. 604 taxa de mutação, p. 606	diversidade nucleotídica, p. 604 taxa de mutação, p. 606	diversidade nucleotídica, p. 604 taxa de mutação, p. 606	diversidade nucleotídica, p. 604 taxa de mutação, p. 606
19. Herança de Traços Complexos						experimento de seleção a longo prazo, p. 650	experimento de seleção a longo prazo, p. 650
20. Evolução de Genes e Traços	experimento de seleção, p. 680 evolução da resistência a antibiótico, p. 681					distinções de espécies, pp. 683-684 evolução da regulação gênica e mancha na asa, pp. 686-688	

(Fotos, da esquerda para a direita: Biophoto Associates/Science Photo Library/Science Source; SciMAT/Science Source; Courtesy of Anthony Griffiths/Olivera Gavric; Darwin Dale/Science Source; Sinclair Stammers/Science Source; Eye of Science/Science Source; R. L. Brinster, School of Veterinary Medicine, University of Pennsylvania.)